

GenCore version 5.1.1.8
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 31, 2006, 10:18:03 ; Search time 17609 Seconds
(without alignments)
11497.390 Million cell updates/sec

Title: US-10-501-930-2
Perfect score: 3166
Sequence: 1 cggagcggcagtcggtgcc.....gctattagcaaaaaaaaaa 3166

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 6366136 seqs, 31973710525 residues

Total number of hits satisfying chosen parameters: 12722272

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Result No.	Score	Query Match	Length	DB ID	Description
1	3166	100.0	3166	6 AB059554	AB059554 Mus muscu
2	2361	74.6	110000	12 BX322642	BX322642 Mus muscu
3	2361	74.6	110000	12 BX322642	Continuation (3 of
4	2361	74.6	110000	12 BX322642	Continuation (3 of
5	2361	74.6	202001	6 AL772303	AL772303 Mouse DNA
6	1408.4	44.5	236283	12 AC120773	AC120773 Rattus no
7	1408.4	44.5	264245	12 AC096804	AC096804 Rattus no
8	1028	32.5	1197	6 AJ699423	AJ699423 Rattus no
9	861.8	27.2	1234	5 AJ621583	AJ621583 Homo sapi
10	857.8	27.1	1197	5 AJ697663	AJ697663 Pan trogl
11	757	23.9	1197	14 AJ868431	AJ868431 Bos tauru
12	559.8	17.7	1149	2 CQ736679	CQ736679 Sequence
13	470.4	14.9	151289	5 AL158164	AL158164 Human DNA
14	467.2	14.8	1197	11 AJ699424	AJ699424 Gallus ga
15	456.8	14.4	1520	11 CR352448	CR352448 Gallus ga
16	381	12.0	112099	12 AC143457	AC143457 Macaca mu
17	257.8	8.1	493	7 BV211170	BV211170 SIAT8F_5
18	256.6	8.1	1334	11 AJ704564	AJ704564 Gallus ga

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

19	248.2	7.8	1140	11	AJ715548
20	247.6	7.8	1131	6	AJ63422
21	244	7.7	1131	5	AJ67662
22	244	7.7	1131	5	CR47037
23	244	7.7	1878	2	CQ71379
24	244	7.7	1953	5	BC13910
25	244	7.7	1953	5	BC13911
26	242.8	7.7	1881	6	BC04855
27	242.4	7.7	2594	5	AK058270
28	241.2	7.6	1854	6	MMA38T
29	239.2	7.6	1899	5	HSU1641
30	237.8	7.5	2135	5	AK04273
31	234.4	7.4	1125	14	AJ58432
32	231.6	7.3	2094	5	AK10126
33	228.8	7.2	1032	11	AJ15546
34	228.2	7.2	1137	11	AJ71609
35	224.4	7.1	1071	2	AX41522
36	224.4	7.1	1630	5	HUM03G
37	224.4	7.1	1704	5	HUM03S
38	224.4	7.1	1755	2	CQ71891
39	224.4	7.1	2117	2	AK35634
40	224.4	7.1	2117	5	HSG3S
41	222.8	7.0	1650	5	HUM3SIATR
42	221.2	7.0	1071	5	AJ67658
43	220.4	7.0	1116	11	AJ705092
44	220.4	7.0	1140	11	AJ15547
45	211.6	6.7	2381	6	BC04821

ALIGNMENTS

RESULT 1	AB059554	166 bp	mRNA	linear	ROD 27-APR-2005
LOCUS	Mus musculus ST8Sia VI	mRNA for alpha 2,8-sialyltransferase,			
DEFINITION	complete cds.				
ACCESSION	AB059554				
VERSION	AB059554.1	GI:21668466			
KEYWORDS	Mus musculus (house mouse)				
SOURCE	Mus musculus				
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muroidea; Muridae; Mus;				
REFERENCE	1	Takahashi, S., Ishida, H., K. Inazu, T., Ando, T., Ishida, H., Kiso, M., Tsuji, S. and Tsujimoto, M.			
AUTHORS	Molecular cloning and expression of a sixth type of alpha 2,8-sialyltransferase (ST8Sia VI) that sialylates O-glycans				
TITLE	J. Biol. Chem. 277 (27), 24030-24038 (2002)				
JOURNAL	11980897				
PUBMED	2 (bases 1 to 3166)				
REFERENCE	2 (bases 1 to 3166)				
AUTHORS	Takahashi, S.				
TITLE	Direct Submission				
JOURNAL	Submitted (07-APR-2001) Shou Takashima, RIKEN, Cellular Biochemistry Laboratory; 2-1 Hirosawa, Wako, Saitama, 351-0198, Japan (E-mail: staka@riken.jp, Tel:81-48-462-1111(ex.3424), Fax:81-48-462-4670)				
FEATURES	Location/Qualifiers				
source	1. 3166				
	/organism="Mus musculus"				
	/mol_type="mRNA"				
	/db_xref="taxon 10090"				
	/country="Japan"				
gene	1. 3166				
	/gene="ST8Sia VI"				
	/77..1273				
CDS	/codon_start=1				
	/product="alpha 2,8-sialyltransferase"				
	/protein_id="BA01265.1"				
	/db_xref="GI:21668467"				


```
1981 GTTCAACATCATCTGCTTCTCACATGTTTATCCATAAAGTTACTCATCTGATTTTATT 2040
1981 GTTCAACATCATCTGCTTCTCACATGTTTATCCATAAAGTTACTCATCTGATTTTATT 2040
2041 TAAATAGTGAACATCTACTTGTATATCAGACCCGAGGACCATCTCTCCATTTGGAGAAATG 2100
2041 TAAATAGTGAACATCTACTTGTATATCAGACCCGAGGACCATCTCTCCATTTGGAGAAATG 2100
2101 AAGATATTTGTCAGTGGCAGAAAAGCAGGTGTGTGCTCAATTAATGATAAGATACCAAGC 2160
2101 AAGATATTTGTCAGTGGCAGAAAAGCAGGTGTGTGCTCAATTAATGATAAGATACCAAGC 2160
2161 ATCATCATGCCAGTATTAACACACAGTGTGAAAGGATCATAGACAGGGGTGTTAAATCT 2220
2161 ATCATCATGCCAGTATTAACACACAGTGTGAAAGGATCATAGACAGGGGTGTTAAATCT 2220
2221 GATCCCATGAGTAATAAATCTTCACTGTACTTATTTTTCAGGGAAGAGTTAATTTTCACAATTA 2280
2221 GATCCCATGAGTAATAAATCTTCACTGTACTTATTTTTCAGGGAAGAGTTAATTTTCACAATTA 2280
2281 AACTAGTAATGAACCAATTTCTTAGGACATTAAGTGAATCTTCACTAGTAAGAAAGGAA 2340
2281 AACTAGTAATGAACCAATTTCTTAGGACATTAAGTGAATCTTCACTAGTAAGAAAGGAA 2340
2341 CAGCAGGAGAAAGCTGTTCGCTTCTGATTAACCCAAATGAGCATGCTGGAAGGAGGT 2400
2341 CAGCAGGAGAAAGCTGTTCGCTTCTGATTAACCCAAATGAGCATGCTGGAAGGAGGT 2400
2401 TGTGAGGCTACGCTAAAACCTCTCGTAGGAGAGAGTACAGTGCATGAGTGTGCGCGCT 2460
2401 TGTGAGGCTACGCTAAAACCTCTCGTAGGAGAGAGTACAGTGCATGAGTGTGCGCGCT 2460
2461 TTTGTCACACTCTGAGGGTGAGTAAATTCAGAGCCAAATCACAATCAAGGATGACAC 2520
2461 TTTGTCACACTCTGAGGGTGAGTAAATTCAGAGCCAAATCACAATCAAGGATGACAC 2520
2521 ACCTAACTCATCACTTCAGGGGGAGAGTAAGTCTTCATGAGAAATACACTCATAGCT 2580
2521 ACCTAACTCATCACTTCAGGGGGAGAGTAAGTCTTCATGAGAAATACACTCATAGCT 2580
2581 AAGCATCAGTTTTCAGTAAAAATTTGAGTAGATGTTAAATATGAACATTTTATACCTCTTA 2640
2581 AAGCATCAGTTTTCAGTAAAAATTTGAGTAGATGTTAAATATGAACATTTTATACCTCTTA 2640
2641 CTAATGTCACCAAGCACTTTTAAATGTAAGCAATTAATTAATTAAGTACTTGAACAT 2700
2641 CTAATGTCACCAAGCACTTTTAAATGTAAGCAATTAATTAATTAAGTACTTGAACAT 2700
2701 AATGCTTATGCTGTATATCTGTTTCATCCATCGATTTTCCCAAAAGTAAGGACATAG 2760
2701 AATGCTTATGCTGTATATCTGTTTCATCCATCGATTTTCCCAAAAGTAAGGACATAG 2760
2761 GAGATGAGGCTACATGCCAAGAAAATATAAAATTTTACTCTTTAAATCTTACTTGAGCC 2820
2761 GAGATGAGGCTACATGCCAAGAAAATATAAAATTTTACTCTTTAAATCTTACTTGAGCC 2820
2821 AGCTTGTTGTTTATCAAGTCTTTTGAAGAGACAGACCCCTGTGAATTTCTTCAATCTG 2880
2821 AGCTTGTTGTTTATCAAGTCTTTTGAAGAGACAGACCCCTGTGAATTTCTTCAATCTG 2880
2881 ATACAGTGTACCTTGTATTAAATTTGAATTTGAATTTGTTTCAAGTTTACATCTTTTCAT 2940
2881 ATACAGTGTACCTTGTATTAAATTTGAATTTGAATTTGTTTCAAGTTTACATCTTTTCAT 2940
2941 TCTTTTATAGCAATCAAAAGTATTAGCTTCAGAAAATTTATCAGAAGTTTCATATATAAT 3000
2941 TCTTTTATAGCAATCAAAAGTATTAGCTTCAGAAAATTTATCAGAAGTTTCATATATAAT 3000
3001 ATTTTGCAGAGGGTAAAGGCTTTTGTGTAATAATAAAATTTTATTTTCTTCTG 3060
3001 ATTTTGCAGAGGGTAAAGGCTTTTGTGTAATAATAAAATTTTATTTTCTTCTG 3060
3061 ATGAATAGAGGCTTTTATGCTGCTGCTAATGAACCTAATTAGCTTTAAATATCTCTCT 3120
```

```
Db 3061 ATGAATAGAGGCTTTTATGCTGCTGCTAATGAACCTAATTAGCTTTAAATATCTCTCT 3120
Qy 3121 AGCAACATTTGGTTCACGTTTCAATCATGCTATTAGCAACCTAATTAATATCTCTCT 3166
Db 3121 AGCAACATTTGGTTCACGTTTCAATCATGCTATTAGCAACCTAATTAATATCTCTCT 3166

RESULT 2
BX322642 0/c
WPCOMMENT
Sequence split into 5 fragments
Fragment Name Begin End
BX322642_0 1 11000
BX322642_1 10001 21000
BX322642_2 20001 31000
BX322642_3 30001 41000
BX322642_4 40001 44002
LOCUS BX322642 44002 bp DNA linear HTG 03-FEB-2006
DEFINITION Mus musculus chromosome 1 clone RP24-36309, WORKING DRAFT SEQUENCE,
8 unordered pieces.
ACCESSION BX322642
VERSION BX322642.14 GI:86476683
KEYWORDS HTG; HTGS PHASE1; HTGS ACTIVEFIN; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE Mus musculus
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 440802)
AUTHORS Lloyd,D.
TITLE Direct Submission
JOURNAL Submitted (02-FEB-2006) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries: vegasanger.ac.uk
COMMENT Cloned requests: clonerequest@sanger.ac.uk
On Feb 4, 2006 this sequence version replaced gi:76666521.
----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: vegasanger.ac.uk
----- Project Information
Center project name: zKp029
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Chemistry: Dye-terminator; 100% of reads
Consensus quality: 27287; bases at least Q40
Consensus quality: 27462; bases at least Q30
Consensus quality: 27539; bases at least Q20
Insert size: 440102; sum-of-contigs
Quality coverage: 6.40x in Q20 bases; agarose-fp
coverage: 16.59x in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 8 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 163997: contig of 163997 bp in length
* 163998 164097: gap of 100 bp
* 164098 166408: contig of 2311 bp in length
* 166409 166509: gap of 101 bp
* 166510 170049: contig of 3540 bp in length
* 170050 170150: gap of 101 bp
* 170151 214018: contig of 4368 bp in length
* 214019 232782: contig of 100 bp
* 232783 232883: contig of 18664 bp in length
* 232885 258463: contig of 25579 bp in length
```


QY	2410	ACGCTAAACCTCTCGCTAGGAGAGAGTA	CAGTGCATGAGTGTGGCGCTTTTCTCCAC	2469
DB	20083	ACGCTAAACCTCTCGCTAGGAGAGAGTA	CAGTGCATGAGTGTGGCGCTTTTCTCCAC	20024
QY	2470	ACTCGTGAAGGGTGAAGTAATTCAGAGCCAA	TCAATCAAGGATGGACACACCTTAACTC	2529
DB	20023	ACTCGTGAAGGGTGAAGTAATTCAGAGCCAA	TCAATCAAGGATGGACACACCTTAACTC	19964
QY	2530	ATCACTTCAGGGGAGATGAATGCTTTTCATG	AGAAATTTACACTCATAGCTTAAGCATCAG	2589
DB	19963	ATCACTTCAGGGGAGATGAATGCTTTTCATG	AGAAATTTACACTCATAGCTTAAGCATCAG	19904
QY	2590	TTTTGAGTAAATTTGAGTAGATGTTAAATAT	TGAACATTTTATACCTCTTACTAATGTC	2649
DB	19903	TTTTGAGTAAATTTGAGTAGATGTTAAATAT	TGAACATTTTATACCTCTTACTAATGTC	19844
QY	2650	CACCGACACCTTTTAATGTAAGCACATTTAT	TATTAAGTTACTTGTACATTAATGCTTA	2709
DB	19843	CACCGACACCTTTTAATGTAAGCACATTTAT	TATTAAGTTACTTGTACATTAATGCTTA	19784
QY	2710	TGCTGTATATTCGTTTCATCATGATGATTTT	CCCAAAAGTAAGCATAGAGATGAGG	2769
DB	19783	TGCTGTATATTCGTTTCATCATGATGATTTT	CCCAAAAGTAAGCATAGAGATGAGG	19724
QY	2770	CCTACATGCCAAGAAACTATAATTTTACTCTT	TAAATTTTACTTGTAGCCAGCTTGTG	2829
DB	19723	CCTACATGCCAAGAAACTATAATTTTACTCTT	TAAATTTTACTTGTAGCCAGCTTGTG	19664
QY	2830	TTTTATCAAGTGCCTTTTGAAGAGACAGCAC	CCCTGTGAAATTTCTTCAATCTGATACAGT	2889
DB	19663	TTTTATCAAGTGCCTTTTGAAGAGACAGCAC	CCCTGTGAAATTTCTTCAATCTGATACAGT	19604
QY	2890	CACCTTGATTAATTAATTTGATTTGATTTT	CAAGTTTACATCTCTTCAATCTTTTATA	2949
DB	19603	CACCTTGATTAATTAATTTGATTTGATTTT	CAAGTTTACATCTCTTCAATCTTTTATA	19544
QY	2950	GCAATCAAGCGTATTAGCTTCAGAAATTTAT	CAGAACTTCATATATAAATTTTGC	3009
DB	19543	GCAATCAAGCGTATTAGCTTCAGAAATTTAT	CAGAACTTCATATATAAATTTTGC	19484
QY	3010	AGGGTAAAGCGCTTTTGTGTTAAATAAATA	ATTAATTTTCTTCTGATGAATAGA	3069
DB	19483	AGGGTAAAGCGCTTTTGTGTTAAATAAATA	ATTAATTTTCTTCTGATGAATAGA	19424
QY	3070	GGCTCTTTTATGCTGCTGAATGAACCTAAT	TAGCTTTAAATATCTCCAGCAACAT	3129
DB	19423	GGCTCTTTTATGCTGCTGAATGAACCTAAT	TAGCTTTAAATATCTCCAGCAACAT	19364
QY	3130	GGTCACGTTTCAATCATGCTATTAGCAAAAA	AAAAA 3166	
DB	19363	GGTCACGTTTCAATCATGCTATTAGCAAAAA	AAAAA 19327	
RESULT 3				
BX322642 2				
WPCOMMENT				
Sequence split into 5 fragments LOCUS BX322642 Accession BX322642				
Fragment Name Begin End				
BX322642_0 1 110000				
BX322642_1 100001 210000				
BX322642_2 200001 310000				
BX322642_3 300001 410000				
BX322642_4 400001 440802				
Continuation (3 of 5) of BX322642 from base 200001 (BX322642 Mus musculus chromosome 2)				
Query Match 74.6%; Score 2361; DB 12; Length 110000;				
Best Local Similarity 99.6%; Pred. No. 0;				
Matches 2367; Conservative 0; Mismatches 10; Indels 0; Gaps 0;				
QY	790	CATTATACCCCTGAAGTACCAAGATTTTGAAG	GAGAGAAACACAGATTTTGGAGGACAT	849
DB	36760	CTTTTAAATTCGGCAGGTACCAAGATTTTGA	GAGAGAAACACAGATTTTGGAGGACAT	36819

QY	850	CTCCACCTATGAGATGATCTTC	TCCTCTCCAGCATTTTCTCTATCGGCGCAACACAG	909
DB	36820	CTCCACCTATGAGATGATCTTC	TCCTCTCCAGCATTTTCTCTATCGGCGCAACACAG	36879
QY	910	CATCTCTTTTAAAGTCTACCAA	CACCTCAAAGAGTCAAAAATAGGCGCAAAAGTTCTCTT	969
DB	36880	CATCTCTTTTAAAGTCTACCAA	CACCTCAAAGAGTCAAAAATAGGCGCAAAAGTTCTCTT	36939
QY	970	CTTCATCCAGGATACCTGAGA	ACCTGCTCTTTTCTGGAGAACTAAAGGGGTGACTGC	1029
DB	36940	CTTCATCCAGGATACCTGAGA	ACCTGCTCTTTTCTGGAGAACTAAAGGGGTGACTGC	36999
QY	1030	ATACCGCTGTGTCACAGGCTTG	TGATTGCAAGTGTGCTGTGGAACTGTGTGAAAACT	1089
DB	37000	ATACCGCTGTGTCACAGGCTTG	TGATTGCAAGTGTGCTGTGGAACTGTGTGAAAACT	37059
QY	1090	GAAGCTCTACGGATTTCTGGCCT	TCTCTAAGACTATCGAAGACACCCCACTCAGTCACCA	1149
DB	37060	GAAGCTCTACGGATTTCTGGCCT	TCTCTAAGACTATCGAAGACACCCCACTCAGTCACCA	37119
QY	1150	CTACTATGATTAACATGTTACCT	TAGCATGGTTTCCACAGATGCTTAAAGAAATACAGCA	1209
DB	37120	CTACTATGATTAACATGTTACCT	TAGCATGGTTTCCACAGATGCTTAAAGAAATACAGCA	37179
QY	1210	AATGCTCCAGCTCCATATGAGA	GAATCTCAAACTGCAATTCAGCAAAATGTGAAACGCG	1269
DB	37180	AATGCTCCAGCTCCATATGAGA	GAATCTCAAACTGCAATTCAGCAAAATGTGAAACGCG	37239
QY	1270	TTAAGCTTTCTTAGAAGGAGAA	TAATTTGAGGAGTGGATGGATGTGTCAAGCATCTTC	1329
DB	37240	TTAAGCTTTCTTAGAAGGAGAA	TAATTTGAGGAGTGGATGGATGTGTCAAGCATCTTC	37299
QY	1330	CAAAAAGCAATAGAAAGGCG	CAGAGAAAGCATGAATTAAGAAAGGCGCTCTCCACATT	1389
DB	37300	CAAAAAGCAATAGAAAGGCG	CAGAGAAAGCATGAATTAAGAAAGGCGCTCTCCACATT	37359
QY	1390	GTCTAGACCAAGCCACCGCC	CCACTCACTTTGCGAGCTCCAGAGTCACTCATCTTC	1449
DB	37360	GTCTAGACCAAGCCACCGCC	CCACTCACTTTGCGAGCTCCAGAGTCACTCATCTTC	37419
QY	1450	ACCTTCAACGTTCTTTCTCTGA	TAATAGAGACCAAAACATCAGACTTGGATAAGTAAAT	1509
DB	37420	ACCTTCAACGTTCTTTCTCTGA	TAATAGAGACCAAAACATCAGACTTGGATAAGTAAAT	37479
QY	1510	GAGATAATTTTCAAAATCATCA	AGAAATTTGATTTGAGCCAGGGTCTCTCAGAAATGCTTC	1569
DB	37480	GAGATAATTTTCAAAATCATCA	AGAAATTTGATTTGAGCCAGGGTCTCTCAGAAATGCTTC	37539
QY	1570	CTTGTTCTTATCCATGATAGCC	TTCCACCTTTTATCAGAGTGGTAAATGAAACTGTGCA	1629
DB	37540	CTTGTTCTTATCCATGATAGCC	TTCCACCTTTTATCAGAGTGGTAAATGAAACTGTGCA	37599
QY	1630	TTGTGCCAAAGACCCCTTTCTGA	BAGAAATGTCTGAATCATCGCGGAGTTTTTCACACA	1689
DB	37600	TTGTGCCAAAGACCCCTTTCTGA	BAGAAATGTCTGAATCATCGCGGAGTTTTTCACACA	37659
QY	1690	GCTCTTCTTTTATAAATAAATC	TTCCCATTTCTCCCTCTAGTAGAGTACAGAAACAAAA	1749
DB	37660	GCTCTTCTTTTATAAATAAATC	TTCCCATTTCTCCCTCTAGTAGAGTACAGAAACAAAA	37719
QY	1750	TACCTTGTAGATTCAGGAAGA	TAGTCTTTTCTTTTACTTAGCAATGTGCTGCTTCTGATT	1809
DB	37720	TACCTTGTAGATTCAGGAAGA	TAGTCTTTTCTTTTACTTAGCAATGTGCTGCTTCTGATT	37779
QY	1810	CAGTTCCCTGTGACATTAAGC	GGTTGGGGTTTGGTTGGATTTGGGGGCTTTCTTCA	1869
DB	37780	CAGTTCCCTGTGACATTAAGC	GGTTGGGGTTTGGTTGGATTTGGGGGCTTTCTTCA	37839
QY	1870	CTTCTTTTGTCTATTAATTTCT	ACCTTTATCAGTTTGTATTCAGAGTCTCTGCTTTGGG	1929
DB	37840	CTTCTTTTGTCTATTAATTTCT	ACCTTTATCAGTTTGTATTCAGAGTCTCTGCTTTGGG	37899
QY	1930	ATTCTGCAATTTCTCTCCCACT	TACAGGATCAACTCAATGACATATAAGTAGTAGTCAACA	1989

Db	37900	ATTCTGCAATTTCTCTCTCCAGTACAGATCACTCAATGACATTAAGTAGTTCACAA	37959
Qy	1990	TCCATTGCTTCTCATGATTTTATCCATAAAGTACTCATCTGATTTTATTAATAATAGT	2049
Db	37960	TCCATTGCTTCTCATGATTTTATCCATAAAGTACTCATCTGATTTTATTAATAATAGT	38019
Qy	2050	GAAATCTTCTGATATATCAGACCCGAGGAGCACCCTCCATTTGGAGATATGAGATATG	2109
Db	38020	GAAATCTTCTGATATATCAGACCCGAGGAGCACCCTCCATTTGGAGATATGAGATATG	38079
Qy	2110	TCACCTGGCAGAAAAGCAGGTGTGTGCCATTAATTGATAAGATACCAAGCATCATCATG	2169
Db	38080	TCACCTGGCAGAAAAGCAGGTGTGTGCCATTAATTGATAAGATACCAAGCATCATCATG	38139
Qy	2170	CCAGTTATGAACACAGTGTCTGAAGGATCATAGACAGGGGTGGTTAAATCTGATCCAGT	2229
Db	38140	CCAGTTATGAACACAGTGTCTGAAGGATCATAGACAGGGGTGGTTAAATCTGATCCAGT	38199
Qy	2230	AGATTAACCTTCACTGATCTATTTTCAAGGAGAGTTAAATTTCAATTAATACTAGTAA	2289
Db	38200	AGATTAACCTTCACTGATCTATTTTCAAGGAGAGTTAAATTTCAATTAATACTAGTAA	38259
Qy	2290	ATGAACCAATTTCTTAGGCACATTAAGTGGATTTCTGAGTAAAGAAAGGAAACAGCAGGAG	2349
Db	38260	ATGAACCAATTTCTTAGGCACATTAAGTGGATTTCTGAGTAAAGAAAGGAAACAGCAGGAG	38319
Qy	2350	AAAGCTGTTCGCTTGGTTCTGATTTACCCAAATGAGCATGCTGGAAAGAGGTTGTGAGCT	2409
Db	38320	AAAGCTGTTCGCTTGGTTCTGATTTACCCAAATGAGCATGCTGGAAAGAGGTTGTGAGCT	38379
Qy	2410	ACGCTAAACCTCTCGCTAGGAGGAGTACAGTGCATGAGTGGCGGCTTTTGTCCAC	2469
Db	38380	ACGCTAAACCTCTCGCTAGGAGGAGTACAGTGCATGAGTGGCGGCTTTTGTCCAC	38439
Qy	2470	ACTCGTGAAGGGTGAGTAAATTCAGAGCCAAATCACATCAAGGATGGACACACCTAACTC	2529
Db	38440	ACTCGTGAAGGGTGAGTAAATTCAGAGCCAAATCACATCAAGGATGGACACACCTAACTC	38499
Qy	2530	ATCACTTCAGGGGAGATGAATGCTTTCATGAGAAAATTAACCTCATTAAGCTAAGCATCAG	2589
Db	38500	ATCACTTCAGGGGAGATGAATGCTTTCATGAGAAAATTAACCTCATTAAGCTAAGCATCAG	38559
Qy	2590	TTTTGAGTAAATTTGAGTACATTTAAATATGACATTTTATACCTCTTACTAATCTCC	2649
Db	38560	TTTTGAGTAAATTTGAGTACATTTAAATATGACATTTTATACCTCTTACTAATCTCC	38619
Qy	2650	CACCGACACCTTTTAAATGTAAGCACATTTATTTAATTAAGTTTACTTGACATTAATGCTTA	2709
Db	38620	CACCGACACCTTTTAAATGTAAGCACATTTATTTAATTAAGTTTACTTGACATTAATGCTTA	38679
Qy	2710	TGCTGTATATTTGTTTATCTCATCGATTTTCCAAAAGGTAAGAGATAGGATGAGG	2769
Db	38680	TGCTGTATATTTGTTTATCTCATCGATTTTCCAAAAGGTAAGAGATAGGATGAGG	38739
Qy	2770	CCTACATGCCAAGAAACTATTAATTTTACCTTTTAAATTTTACTTCTGAGCCAGCTTGTG	2829
Db	38740	CCTACATGCCAAGAAACTATTAATTTTACCTTTTAAATTTTACTTCTGAGCCAGCTTGTG	38799
Qy	2830	TTTTATCAAGTGTCTTTTGAAGAGACAGCACCTGTGAAATTTCTTCAATTCGATACAGTGT	2889
Db	38800	TTTTATCAAGTGTCTTTTGAAGAGACAGCACCTGTGAAATTTCTTCAATTCGATACAGTGT	38859
Qy	2890	CACCTTGATTTTAAATGTAATGTTGTTTCAAGTTTAAATCTCTTTTCAATCTTTTATA	2949
Db	38860	CACCTTGATTTTAAATGTAATGTTGTTTCAAGTTTAAATCTCTTTTCAATCTTTTATA	38919
Qy	2950	GCAAAATCAAGGATTAAGTTCAGAAATTTATCAGAGGTTTCATATATAAATTTTTCGAA	3009
Db	38920	GCAAAATCAAGGATTAAGTTCAGAAATTTATCAGAGGTTTCATATATAAATTTTTCGAA	38979
Qy	3010	AGGTAATAAGGCTTTTGTGTTTAAATTAATAAATTTTATTTTCTTCTGATGAATAGA	3069

Db	38980	AGGTAATAAGGCTTTTGTGTTTAAATAATAAATTTTATTTTCTTCTGATGAATAGA	39039
Qy	3070	GGCTCTTTTATGCTGTGCTTAAATTAATTAAGCTTTTAAATTTATCTCTAGCAATTT	3129
Db	39040	GGCTCTTTTATGCTGTGCTTAAATTAATTAAGCTTTTAAATTTATCTCTAGCAATTT	39099
Qy	3130	GGTCACGCTTTCAATCATGCTATTAGCAAAAAAATAAATAA 3166	
Db	39100	GGTCACGCTTTCAATCATGCTATTAGCAAAAAAATAAATAA 39136	
RESULT 4			
BX322642 2/c			
WPCOMMENT			
Sequence split into 5 fragments			
Fragment Name Begin End			
BX322642_0 1 110000			
BX322642_1 100001 210000			
BX322642_2 200001 310000			
BX322642_3 300001 410000			
BX322642_4 400001 440002			
Continuation (3 of 5) of BX322642 from base 200001 (BX322642 Mus musculus chromosome 2)			
Query Match 74.6%; Score 2361; DB 12; Length 110000;			
Best Local Similarity 99.6%; Prob. No. 0;			
Matches 2367; Conservative 0; Mismatches 10; Indels 0; Gaps 0;			
Qy	790	CATTATTAACCTGAGTACAGATTTGAAGGAGAGAAAGACAGATTTTGGAGGACAT	849
Db	98506	CTTTTAATTCGGCAGGTACCAATTTGAAGGAGAGAAAGACAGATTTTGGAGGACAT	98447
Qy	850	CTCCACCTATGGAGATGATTCCTCTCTGCCAGCATTTTCTCTATCGGGCAACACAGG	909
Db	98446	CTCCACCTATGGAGATGATTCCTCTCTGCCAGCATTTTCTCTATCGGGCAACACAGG	98387
Qy	910	CATCTCTTTTAAAGTCTACCAATTAATAAGAGTCAAAAATAGGCAAAAGGTTCTCTT	969
Db	98386	CATCTCTTTTAAAGTCTACCAATTAATAAGAGTCAAAAATAGGCAAAAGGTTCTCTT	98327
Qy	970	CTTTCATCCAGGATCTCTGAGATCTCTGAGATCTCTCTCTCTCTCTCTCTCTCTCT	1029
Db	98326	CTTTCATCCAGGATCTCTGAGATCTCTGAGATCTCTCTCTCTCTCTCTCTCTCTCT	98267
Qy	1030	ATACCGCTTCTCCACAGGCTTGAATTCAGAGTGTGCTGTGGAACTGTGTGAAAACGT	1089
Db	98266	ATACCGCTTCTCCACAGGCTTGAATTCAGAGTGTGCTGTGGAACTGTGTGAAAACGT	98207
Qy	1090	GAAGCTCTACGGATTTCTGGCTTCTCTAAGACTATCGAAGACACCCCACTCAGTCACCA	1149
Db	98206	GAAGCTCTACGGATTTCTGGCTTCTCTAAGACTATCGAAGACACCCCACTCAGTCACCA	98147
Qy	1150	CTACTATGATAACATGTTTACCTAGCATGTTTCCACAGATGCTTAAGAAATACAGCCA	1209
Db	98146	CTACTATGATAACATGTTTACCTAGCATGTTTCCACAGATGCTTAAGAAATACAGCCA	98087
Qy	1210	AATGCTCCAGCTCCATATGAGATGAATCTCAAACTCAATTCAGCAAAATGTGAAAACGG	1269
Db	98086	AATGCTCCAGCTCCATATGAGATGAATCTCAAACTCAATTCAGCAAAATGTGAAAACGG	98027
Qy	1270	TTAACGCTTTCTTGAAGGAGAAATTAATTCAGGAGGTGAGTGTGTCTACAGCATCTC	1329
Db	98026	TTAACGCTTTCTTGAAGGAGAAATTAATTCAGGAGGTGAGTGTGTCTACAGCATCTC	97967
Qy	1330	CAAAAAGCCCAATAGAAAGAGGCTTAGAAGAGCATGAATTTACAAAGGGCTCTCCCACTT	1389
Db	97966	CAAAAAGCCCAATAGAAAGAGGCTTAGAAGAGCATGAATTTACAAAGGGCTCTCCCACTT	97907
Qy	1390	GTCTAGACCAAAAGCCACCCGCCCTCACTCTTTGACGCTCCACAGTCACTCATCTCTC	1449
Db	97906	GTCTAGACCAAAAGCCACCCGCCCTCACTCTTTGACGCTCCACAGTCACTCATCTCTC	97847
Qy	1450	ACCTTCAACGCTTCTTCTCTGACATATAGAGACCAAAATCATCAGACTTTGGATTAAGTAAAT	1509


```
----- Project Information
Center project name: GZJM
Center clone name: CH230-65J6
----- Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 221191 bases at least Q40
Consensus quality: 224201 bases at least Q30
Consensus quality: 225744 bases at least Q20
Estimated insert size: 228488; sum-of-contigs estimation
Quality coverage: 7x in Q20 bases; sum-of-contigs estimation
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 2 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 224842: contig of 224842 bp in length
* 224843 224942: gap of unknown length
* 224943 236283: contig of 11341 bp in length.
FEATURES             Location/Qualifiers
     source            1..236283
                        /organism="Rattus norvegicus"
                        /mol_type="genomic DNA"
                        /db_xref="taxon:10116"
                        /clone="CH230-65J6"
     misc_feature      37445..38277
                        /note="clone boundary"
                        clone_end:17
                        site:
     gap               end_sequence:BH351867"
                        224843..224942
                        /estimated_length=unknown

ORIGIN
Query Match          44.5%; Score 1408.4; DB 12; Length 236283;
Best Local Similarity 80.4%; Pred. No. 0;
Matches 1945; Conservative 0; Mismatches 391; Indels 82; Gaps 22;

Qy 806 TACGAGATTTCAAGGAGAGAGACACAGTTTTTTGGAGGACATCTCCACCTATCGAGT 865
Db 57567 TACAGAGATTTGAAGGAGAGAGACACGGTTTCTGGAGGATATTTCCGCCCTATCGGGAC 57626

Qy 866 GCATTCCTCTCTCCAGCATTTTCTATCGGGCAACACAGGCATCTCTTTAAAGTC 925
Db 57627 GCATTCCTCTCTTCCAGCCTTTTCTATCGGGTCAACACCGGCATCTCTTTAAAGTC 57686

Qy 926 TACCAACACCTCAAGAGTCAAAATATGAGGCAAAAGTTCTCTTCCATCCCGAGTAC 985
Db 57687 TACCAACACCTCAAGAGTCTAAGTGAGGCAAAAGTTCTCTTCCACCCAGGATAC 57746

Qy 986 CTGAGACACCTCGCTCTTTCTGGAGAACTAAGGGGTGACTGCATACGGTTGTCACA 1045
Db 57747 CTGAGACACCTCGCTCTTTCTGGAGAACTAAGGGGTGACTGCCTTGTCCACG 57806

Qy 1046 GGCCTTGATTTGCAAGTGTCCGTGTGGAATCTGTGTGAAACGCTGAAGCTCTACCGATTC 1105
Db 57807 GGCCTTGATTTGCAAGTGTCCGTGTGGAATCTGTGTGAAAGTGCACGCTCTACCGATTC 57866

Qy 1106 TGGCCTTTCTTAAGATATGAAGACACCCCACTCAGTCACTATCTATGATTAACATG 1165
Db 57867 TGGCCTTTCTTAAGACCGTCGAAGAAATTCCTCTGAGCACCACTACTATGACAAACAG 57926

Qy 1166 TTACTTAAGCATGGTTTCCACAGATGCTTAAGATACAGCAATCTCCAGCTCCAT 1225
Db 57927 CTACCTTAAGCATGGTTTCCACAGATGCTTAAGATACAGCAATCTCCAGCTCCAT 57986

Qy 1226 ATGAGAGGAATCCTCAACTGCAATTCAGCAATGTGAAACGGCTTAAAGCTTCTTAGAA 1285
Db 1285 ATGAGAGGAATCCTCAACTGCAATTCAGCAATGTGAAACGGCTTAAAGCTTCTTAGAA 1285
```

```
Db 57987 ATGAAGGATCCTCAAACTGC GTTCAGCAAAATGTGAAGCGCTTAGCGTTTCTTAAAG 58046
Qy 1286 GGAGATAATTTTCAGGAGGTGG GTGGATGTGTACAGCATCTCCAAAAGCAATAGNA 1345
Db 58047 TCGGAACCAATTTTCAGTAGTGG GTGGGTGCATCGCAGTGTCTCC-AAACGTGGATGGAG 58105
Qy 1346 GAAGGCACAGAGAAAGCATGAAT TACAAAGCGCTCTCCCACTTGTCTAGACCAAGC-- 1403
Db 58106 AAAGGCACAGAAATGGA-GGA TGCACAGAGCTCCCGACTGGCTAGACCAAGGCC 58164
Qy 1404 -CACCCGCCCACTCACTTTGAGCCTTCCAGAGTCACTCACTTCTCACTTCAACGTTTC 1462
Db 58165 GCCCCGCCCGCTTACCTTTGAGCCCCCAGAGTCACTCACTTCACTTCAATGTTC 58224
Qy 1463 TTTCTCTGAGATAGAGACCAAAACATCAGACTTGGATAAGTAAATGAGATAATTTTC 1522
Db 58225 TTTCTCTGAGATAGAGGTCACTATATCAGACTCCGACAAAGTAAATGAGATCATGTGTA 58284
Qy 1523 AAATCATCATAGAAATTTGATTTAGCCAGGGTCTCTCAGAATGCTTCTTGTTCCTATCC 1582
Db 58285 AATCGGCATAGGATTTGATCTAGGCAGCAGCTCGAGGAGTGTCTTCTCGTTCCTATTC 58344
Qy 1583 ATGATAGCAATTTCCACCTTTAAGAGTGGTAATGAACCTGTGCAATGTGCCAAAGAC 1642
Db 58345 GTGGTAGCCATTTCCCGCCATTAAGAGTGGTGAAGAACTGTGCAACTGTGCCAAAGAG 58404
Qy 1643 CTTTCTGAGAGAAATGTCTGAATCA-----TGGGCCGAGTTTTTACACACAGCTTTC 1696
Db 58405 CCTTCTGAGAGAAATGTCCCGCATAGAGGTGCACGAGCTCTCCCTCACAGGCCCTTC 58464
Qy 1697 CTTTATAAATAA---ATCCCTTTCATTTCCCTCTCTAGTAGAGTACAGAAACAAATATCC 1753
Db 58465 CTTTATCAATAACCTTTCCACTTCCCTCCCTCCCAATGAGTACCGGAACAAACACC 58524
Qy 1754 CTTGATGATTCAGGAAGAAAGTTTTTTTACTTAGCAATGTGCCCTGTCTCTGATTCAGT 1813
Db 58525 CTTGATGATTCAGGAAGAAACCTTTTCTACTTAGCAATGTGCCCTGTCTCTGACTCGGT 58584

Qy 1814 TCGCTTGTGACATTAAGCTGGGTGGGGTTTTGGTGGATTTGGGGCGTTTCTTCACTTC 1873
Db 58585 TCACTTGTGA-TGTAATCTGGGTGAGGTGTGTGGGATTTGGGGCGTTTCTCGCCTC 58643
Qy 1874 TTTTGTCTATATTTTCTTACCCTATCAGTTTGTATTCGAGCTTCTCTGCTTGGGATTC 1933
Db 58644 TTTTGTCAATATTTTGCCTACCTATCAGTTTGTGTAGAGCTTCTTCTCTGCGGTTTC 58703
Qy 1934 TGCAAT-----CTCTCTCCCACTGACAGATCAACTCAATGACATAA 1976
Db 58704 TCTTCTTTGGGGTTCTGTCTCCCTCCACACACACCGACAGGATCCACTCAATGACATAA 58763
Qy 1977 AGTAGTTCAACATCCATTGCTTTCACATGTT---TTATCCATAAAGTTACTCATCTGA 2033
Db 58764 AGTAGTTCCAGTATCTGTCTACTTTCACAGGTTATCATATCCAGGAAGTTACCCGCCGA 58823
Qy 2034 --TTTTTATTTAAATAGTGAACCTACTTATATCAGACCCGAGGACCATCTCCATTTG 2091
Db 58824 CTTTTTTTTTATAGTAACCTATTGGATATCAGACCTGGGACTG-----TTG 58876
Qy 2092 GAGAAATGAAAGATTTGTCTACCTGAGAAAAGCAGGTGTGTGCCATTAATTTGATAAGT 2151
Db 58877 GAGGATATGAAGAGATTTGTCACTGCGAAGAGCGAGTGTGTACCATTAATTTGATAAGAC 58936
Qy 2152 AC---CAACAACATCATCTGCAGTTATGAACACAGTGTGTGAAGAGATCATAGACAGG 2208
Db 58937 ACGGTAGCAGCGGCCATCGTGAAGTTATGACAGCTGTCTGGAAAGGATCATAGCAGGA 58996
Qy 2209 GTGGTAAATCTGATCCAGTAAATAACTTCACTAGTGTACCTATTTCAGGGGAAGAGTTAA 2268
Db 58997 ACAGTTAAATCTGATTCAGTAAATCAATGCAATATGTCTATTTCAGGGGAAGAGTTGA 59056
Qy 2269 TTTCACAATTTAAACTAGTAAATGAACCAATTTTAGGCACATTAAGTGGATTCAGTGA 2328
Db 59057 TTTCACAATTTAAACTAGTAAATGAACCAATTTCTTTA-GCAAGTTAAGTGGATTCGGGGTA 59115
```

QY	2329	AAAGAAAG-----GGAACAGCAGGAGAAAGCTGTTCGCTTGTGTTCTGATTATCCCAAATGA	2383
Db	59116	AAAAAAAAAATGAAACAGAGAGAGCGGTTTCTCGACTATGATTTTCCAAATGA	59175
QY	2384	GCATGCTGGAAGAGGTTGTGAGGCTAGCTAAACCTCTCGTGGTGGGAGAGATACAGT	2443
Db	59176	GTATGCTGATATGATGCTGTGAGGTTGCTGACACACTCTGTGTAGGAGAGATACAGG	59235
QY	2444	GCATGAGTGTGGCGC-TTTTGTCCACACTCGT---GAAGGTTGAGTAAATTCAGAGCCAA	2499
Db	59236	GCATGGGGTGTAGCTTTTGTCTCCACACTGCTGAAGAGTGTGAGTAAATTCAGAGCCAA	59295
QY	2500	TCATATCA-----AGGATGGACACACCTAACTCATCTTACCTTACGCGGGAGATGA	2549
Db	59296	TCATTTCAAGAAAGGCTTAGGATAGCAGCCCTGACACATTACTCCAGGGAAGATGA	59355
QY	2550	ATGCTTTTCATGAGAAATACACTCATAGCTAAAGCATAGTTTTCAGTAAATTTGAGTA	2609
Db	59356	ACGTTTTTCATGAAATTTGCACTCATGCTGAGCATAGCTTGTGAGTAAATTTGAGTA	59415
QY	2610	GATGTTAAATATGAACATTTTATACCTCTTACTTAATGTCCCAACGACACCTTTTAAATGA	2669
Db	59416	GATGTTAAATATGAACATTTTACACCCCTTAC-----GTCCCAAGGATACCCCTTAAATGA	59471
QY	2670	AGCATATTTATTAATAGTTACTTTGACATTAATAGCTTATGTCTGTATATTC-TGTTCA	2728
Db	59472	AGCATATTTCTTAGTTTATTTGACT--TATTACATATTTATGTGTGATCATTTCTTTTCA	59529
QY	2729	TCCATCGATTTTCCCAAAAGTAAGAGCATAGGATAGGAGCTAGGCTCATGCGCAAGAAACT	2788
Db	59530	TCCAGGAGTTTCCCAAGAACTAGGAGCTAGGAGTGGGTCCTTGTCCCAAGAAACT	59589
QY	2789	ATAAATTTTACTCTTTAATTTCTTACTTGGCCAGCTTGTGTGTATCAAGTCTTTTTTG	2848
Db	59590	ATAAATTTTACTCTTTAATTTCTTACTTGGCCAGCTTGTGTGTATCAAGTCTTTTTG	59648
QY	2849	AAGAGACAGCCCTGTGAATCTTTCATCTGATACAGTGTCACCTGTGATTTTAACTT	2908
Db	59649	AAGGAAGATTACCTGTGAATCTTTCATCTGATACAGTGTCACCTGTGATTTTAACTT	59708
QY	2909	GTAATGTTTTCAGTTTACATCTCTTTTCTTTTCTTTTATAGCAATCAACGATTTAGC	2968
Db	59709	GTAATGTTTTCAGTTTACATCTCTTTTCTTTTCTTTTATAGCAATCAACGATTTAGC	59765
QY	2969	TTCAAGAAATTTATCAGAGTTTCATATATAAATTTTTCGAAGGTTAAAGGCTTTTTTG	3028
Db	59766	TTTGAATATATCAGAGTTTCATATATAAATTTTTCGAAGGTTAAAGGCTTTTTTG	59825
QY	3029	TTAAATAAATAA-----ATTATTTATTTCTTCTGATGATAGAGGCTCTTTTATGCTG	3084
Db	59826	TTAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA	59885
QY	3085	CTGCTAATGAACCTTAATTTAGTTTAAATTAATTTCTCTAGCAATGTCAGTTTCAATC	3144
Db	59886	CTGCTAATGGGTTTAAATTTAGCTTTAAATTAATTTCTCTAGCAATGTCAGTTTCAATC	59945
QY	3145	ATGCTATTAGCAAAAAA 3162	
Db	59946	CCACTATTGCAAGAAA 59963	
RESULT 7	AC096804/C		
LOCUS	Rattus norvegicus clone CH230-44K10, ***	264245 bp	linear
DEFINITION	***, 2 unordered pieces.		HTG 10-MAY-2003
ACCESSION	AC096804		SEQUENCING IN PROGRESS
VERSION	AC096804.6	GI:305211409	
KEYWORDS	HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.		
SOURCE	Rattus norvegicus (Norway rat)		
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		

Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muroidae; Muridae; Murinae; Rattus.	
1 (bases 1 to 264245)	
Murphy, D. Marie, Metzer, Lee, Abramzon, S., Adams, C., Alder, J.,	
Allen, C., Allen, H., Albrecht, S., Amin, A., Anguiano, D.,	
Avodeji, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H.,	
Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F.,	
Biswal, K., Blair, J., Blakenburg, K., Blyth, P., Brown, M.,	
Bryant, N., Buhay, C., Burroughs, P., Burrell, K., Calderon, E.,	
Cardenas, V., Carter, K., Casar, H., Center, A.,	
Chacko, J., Chavez, D., Chen, G., Chen, Y., Chen, Z., Chu, J.,	
Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L.,	
DeVila, M.L., Davis, C., De Franco, C., Ding, Y., Dinh, H., Divya, D.,	
Delgado, O., Denison, S., Draper, H., Dugan-Rocha, S.,	
Egan, A., Escotto, M., Evans, C.A., Falls, T., Fan, G.,	
Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,	
Fraser, C.M., Gabisi, A., Gatta, R., Garcia, A., Garner, T., Garza, M.,	
Gebregiorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, M.,	
Gunaratne, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, J.,	
Harvey, Y., Havlak, P., Hayes, A., Henderson, N., Hernandez, J.,	
Hernandez, R., Hines, S., Hladun, S.L., Hodgson, A., Hognes, M.,	
Hollins, B., Howells, S., Hylk, S., Hume, J., Idlebird, D., Jackson, A.,	
Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A.,	
Karpach, S., Kelly, S., Khan, Z., King, L., Kovar, C., Liu, J.,	
Kowis, C., Kraft, C.L., Lebowitz, H., Levan, J., Lewis, L., Li, Z., Liu, J.,	
Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J.,	
Lorenz, H., Lozano, R.J., Lu, X., Ma, J., Maheshwari, M., Mahindratna, M., Mahmoud, M., Malloy, K., Mangum, A.,	
Mangum, B., Mapua, P., Martin, R., Martinez, E.,	
Mawhinney, S., McLeod, M.P., McNeill, T.Z., Meenen, E.,	
Milosavljevic, A., Miner, J., Minja, E., Montemayor, J., Moore, S.,	
Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Naif, L.,	
Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S.,	
Nwakolam, O., Okwuonu, O., Olarnpunsagoon, A., Pal, S., Parks, K.,	
Pasternak, S., Paul, H., Perez, A., Perez, L., Pfankuch, C.,	
Plommer, P., Poinexter, A., Popovic, D., Primus, E., Pu, L.,	
Puzo, M., Quiroz, J., Racine, B., Reeves, K., Regier, M.A., Reigh, R.,	
Reilly, B., Reilly, M., Reilly, Y., Reuter, M., Richards, S., Riggs, F.,	
Rives, C., Rodkey, T., Roj, A., Rose, M., Rose, R., Ruiz, S.J.,	
Sanders, W., Savary, G., Scher, S., Scott, G., Shatman, S., Shen, H.,	
Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C.D., Smaj, D.,	
Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J.,	
Stearns, M., Strong, R., Stuck, A., Svatek, A., Taber, P., Taylor, C.,	
Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Umami, K.,	
Valas, R., Vera, V., Villalana, D., Waldron, L., Walker, B., Wang, J.,	
Wang, O., Wang, S., Warren, R., Warren, R., Wei, X., White, F.,	
Williams, G., Willson, R., Wleczek, R., Wooden, H., Worley, K.,	
Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V.,	
Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von	
Weinstock, G., and Gibbs, R.A.	
Direct Submission	
2 (bases 1 to 264245)	
Worley, K.C.	
Direct Submission	
Submitted (28-SEP-2001)	
of Molecular and Human Genetics, Baylor College of Medicine, One	
Baylor Plaza, Houston, TX 77030, USA	
3 (bases 1 to 264245)	
Direct Submission	
Rat Genome Sequencing Consortium	
Submitted (10-MAY-2003)	
Human Genome Sequencing Center, Department	
of Molecular and Human Genetics, Baylor College of Medicine, One	
Baylor Plaza, Houston, TX 77030, USA	
On May 10, 2003 this sequence version replaced gi:24819385.	
The sequence in this assembly is a combination of BAC based reads	
and whole genome shotgun sequencing reads assembled using Atlas	
(http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described	
in the feature table below represents a scaffold in the Atlas	
assembly ('a' contig-scaffold). Within each contig-scaffold	
individual sequence contigs are ordered and oriented, and separated	

by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: hgsc-help@bcm.tmc.edu

----- Project Information

Center project name: GFPO

Center clone name: CH230-44K10

----- Summary Statistics

Assembly program: Atlas 3.0;

Consensus quality: 215957 bases at least Q40

Consensus quality: 221425 bases at least Q30

Consensus quality: 225088 bases at least Q20

Estimated insert size: 225712; sum-of-contigs estimation

Quality coverage: 5x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).

* NOTE: This is a 'working draft' sequence. It currently

* consists of 2 contigs. The true order of the pieces

* is not known and their order in this sequence record is

* arbitrary. Gaps between the contigs are represented as

* runs of N, but the exact sizes of the gaps are unknown.

* This record will be updated with the finished sequence

* as soon as it is available and the accession number will

* be preserved.

* 1 262878: contig of 262878 bp in length

* 262879 262878: gap of unknown length

* 262979 264245: contig of 1267 bp in length.

FEATURES

source

1. 264245

/organism="Rattus norvegicus"

/mol_type="genomic DNA"

/db_xref="taxon:10116"

/clone="CH230-44K10"

1. 1967

/note="wgs end extension

clone end:77"

complement(3314..4075)

/note="clone boundary

clone end:77

site:EcoRI

end sequence:BH294791"

80231..81276

/note="wgs contig"

complement(260505..261297)

/note="clone boundary

clone end:Sp6

site:EcoRI

end sequence:BH294793"

262879..262978

/estimated_length=unknown

misc_feature

misc_feature

misc_feature

misc_feature

gap

ORIGIN

Query Match 44.5%; Score 1408.4; DB 12; Length 264245;
 Best Local Similarity 80.4%; Pred. No. 0;
 Matches 1945; Conservative 0; Mismatches 391; Indels 82; Gaps 22;
 QY 806 TACCAGAAATTTGAAGGAGAGAAAGACACAGTTTCTGGAGGACATCTCCACCTATGGAGAT 865
 Db 247502 TACAAGATTTGAAGGAGAGAAAGACACGGTTCTGGAGGATATTTCCGCCTATGGGAC 247443
 QY 866 GCATTCCTCTCTCCAGCATTTTCTTATCGGGCAACAGGATCTCTTTAAAGTC 925
 Db 247442 GCATTCCTCTCTCTCCAGCATTTTCTTATCGGGTCAACACCGGATCTCTTTAAAGTC 247383
 QY 926 TACCAGAACTCAAGAGTCAAAATAGGCGAAAGGTTCTTCTTCCATCCCGAGTAC 985

Db 247382 TACCAGAACTCAAGAGTCTTAAGTGGGCAAAAGGTTCTTCTTCCACCCAGGTAC 247323
 QY 986 CTGAGACACCTCGCTCTTTTCTTCTGAGAACTAAAGGGGTGACTGACATACCGTTGTCCACA 1045
 Db 247322 CTGAGACACCTCGCTCTTTTCTTCTGAGAACTAAAGGGGTGACTGCGTACCGCTGTGCCAG 247263
 QY 1046 GGCCTTGATGATTGCAAGTGTCTGTTGGAACTGTGAAACGCTGAAGCTCTACCGATTTC 1105
 Db 247262 GGCCTTGATGATTGCGGAGCATCGTGTGGAATGTGTGAAATGTACCGCTCTACCGATTTC 247203
 QY 1106 TGGCCTTTTCTTAAGACTATCGTACACCCCACTCAGTCAACCACTACTATATGATAACATG 1165
 Db 247202 TGGCCTTTTCTTAAGACCGTCTGAAATTCCTCTGAGCCACCACTACTATATGATAACATG 247143
 QY 1166 TTACTTAAGCATGTTTCCACCTATGCTTAAGAATACAGCCAAATGCTCAGTCTCAT 1225
 Db 247142 CTACTTAAGCAGCGTTTCCACCTATGCTTAAGAAATACAGCCAAATGCTCAGTCTCAT 247083
 QY 1226 ATGAGAGGATCTCAAACTGCTATTCAGGAAATGTGAACGGCTTAACGTTTCTTAGAA 1285
 Db 247082 ATGAGAGGATCTCAAACTGCTATTCAGGAAATGTGAACGGCTTAACGTTTCTTAGAA 247023
 QY 1286 GGAGAAATATTTCAAGAGGTGGTGTGATGTGTCAAGCATCTCCAAAAAGCCAAATAGAA 1345
 Db 247022 TGGCAACCTTTCAAGTAGTGGTGTGGTGCATCGCAGTGTCTCC-AAAACGTGGATGAG 246964
 QY 1346 GAAGGCAAGAGAAAGCATGAAATACAAAGGCGTCTCCCACTCTGTCTAGACCAAGC-- 1403
 Db 246963 AAAGGCAAGAAATGGA-GGATGCAAGAGCTCCCGGCTTGGCTAGACCAAGGCC 246905
 QY 1404 -CACCGCCGCCCTCACTTGTAGCCTCCACAGTCACTCATCTCACTTCAACGTTTC 1462
 Db 246904 GCCCGCCGCCCTCACTTGTAGCCTCCACAGTCACTCATCTCACTTCAACGTTTC 246845
 QY 1463 TTTCTCTGAGAAATAGAGACCAAAATACATCAGACTTTGGATAAGTAAATGAGATAATTTTC 1522
 Db 246844 TTTCTCTGAGAAATAGAGGTCAATATCAGACTCCGACAAAGTAAATGAGATCATGTTTA 246785
 QY 1523 AAATCATCATAGAAATTTGATTTAGCCAGGCTCTCTAGAAATGCTTCTTGTCTATATCC 1582
 Db 246784 AATCGGCCATAGGATTTGATCTTAGCCAGGCTCTGAGGAGTGTCTTCTCTCTATATTC 246725
 QY 1583 ATGATAGCCATTTCCACCTTTATCAGAGTGTATGAACTGTGCAATTTGTGCCAAAGAC 1642
 Db 246724 GTGTAGCCATTTCCCGCCATTAAGAGTGTGAGAACTGTGCAATTTGTGCCAAAGAC 246665
 QY 1643 CCTTTCTGAAGAGAAATGCTGTAATCA-----TGCGCCGAGTGTATTTTACACACAGCTTTC 1696
 Db 246664 CCTTTCTGAAGAGAAATGCTGTAATCA-----TGCGCCGAGTGTATTTTACACAGCTTTC 246605
 QY 1697 CTTTATAAATAA---ATTCCTTTCATTTCTCTCTAGTAGTAGTACAGAAACAAATATCC 1753
 Db 246604 CTTTATAAATAA---ATTCCTTTCATTTCTCTCTAGTAGTAGTACAGAAACAAATATCC 246545
 QY 1754 CTTGATGATTCAGGAAGAAAAGTTTCTTACTTTAGCAATGTGCTTCTGATTTCACT 1813
 Db 246544 CTTGATGATTCAGGAAGAAAAGTTTCTTACTTTAGCAATGTGCTTCTGATTTCACT 246485
 QY 1814 TCGCTTGTGACATTAAGCTGGTGGGGTTTGGTTGGATTGGGGGGTCTTCTTCACTTTC 1873
 Db 246484 TCACCTTGA-TGTAATCTGGAAGAGTGTGGTGGGATTTGGGGCGTCTTCTCGCCTC 246426
 QY 1874 TTTTGTCTATATTTTCTTACCTATATCAGTTGTATTCGAGCTTCTGCTTTGGGATTC 1933
 Db 246425 TTTGCTCAATACCTTTGCTACCTATATCAGTTGTGTGTAGAGTCTCTCTCTCTGGGGTTC 246366
 QY 1934 TGCAATT-----CTCTCTCCCACTGACAGATCAACTCAATGACATAA 1976
 Db 246365 TCTTCTTTGGGGTCTGCTCTCCCTCCACACACACAGAGATCCACTCAATGACATAA 246306
 QY 1977 AGTAGTTTCAAAATCCATTGCTTTTATCCATGTT---TTATCCATAAAGTTTACTCATCTGA 2033

Db 246305 AGTAGTTCAGTATCTGCTACTCTTCCACAGGTTATCATATCCAGGAAGTTATCCCGCGCGA 246246
 Qy 2034 ---TTTATTTAAATAGTGAACATCTACTTGTATATCAGACCCGAGACCATCTCTCATTTG 2091
 Db 246245 CTTTTTTTTTTAAATAGTAAACATCTATTGGATATCAGACCTGGGACTG-----TTG 246193
 Qy 2092 GAGATATGAGATATTGTCTACTGCGAGAAAAGCAGGTGTGTGCCATTAATTTGATAAGAT 2151
 Db 246192 GAGGATATGAGAGATTGTCTACTGCGAGAGAGCGAGTGTGTACCATTAATTTGATAGAC 246133
 Qy 2152 AC---CACAAGCATCATATCCAGTATATGAAACAGAGTCTGAAAGGATCATAGACAGG 2208
 Db 246132 ACGGTAGCAGGCGCCATCGTACAGTATGACACAGTGTGGAAGGATCATAGCAGGA 246073
 Qy 2209 GTGGTTAAATCTGATCCAGTAAATGAACCTTCACTGTACCTATTTTCAGGAGAGATTAA 2268
 Db 246072 ACAGTTAAATCTGATTCAGTAAATGAACCTTCACTGTACCTATTTTCAGGAGAGATTAA 246013
 Qy 2269 TTTCAAAATTAAGCTAGTAAATGAACCTTCTTAGGCACATTAAGTGGATTCTTGAGTA 2328
 Db 246012 TTTCAAAATTAAGCTAGTAAATGAACCTTCTTAGGCACATTAAGTGGATTCTTGAGTA 245954
 Qy 2329 AAAGAAAG-----GGAACAGCAGGAGAAAGCTGTTCGCTTGTGTTCTGATTACCCAAATGA 2383
 Db 245953 AAAAAAAAAAATGGAACAGAGAGAGAGGGGTTTCTCGACTATGATTTTCCAAATGA 245894
 Qy 2384 GCATGCTGGAAGGAGTTGTGAGGCTAGCTTAAACCTCTCGGTAGGAGAGAGTACAGT 2443
 Db 245893 GTATGCTGATAGTACAGTGTGAGGTTGCTGACACACTCTGTGTTAGGAGAGAGTACAGG 245834
 Qy 2444 GCATGAGTGTGCGGC-TTTTGTCCACACTGTT---GAAGGTGAGTAAATTCAGAGCCAA 2499
 Db 245833 GCATGGGGGTGTAGCTTTTGTCTCACAATGTTGTAAGAGTGTGAGTAAATTCAGAGCCAA 245774
 Qy 2500 TCACATCACA-----AGGATGAGCAGACCTTAATCACTTCCAGGGGAGATGA 2549
 Db 245773 TCACTTCAAGGAGGCTTAGGATAGCAGCGCTGACACACTTACTCCAGGGAAGATGA 245714
 Qy 2550 ATGCTTTCATGAGAAATTAAGTCTAATGAGTAAAGTCAAGTATTTGAGTAAATTTGAGTA 2609
 Db 245713 ACGTTTTCATGAGAAATTTGCACTCAATGGCTAAGCATCAGTCTTGAGTAAATTTGAGTA 245654
 Qy 2610 GATGTTAAATGAACATTTTATATCTCTTACTAATGTCACGACACACCTTTTAAATGTA 2669
 Db 245653 GATGTTAAATGAACATTTTATACCCCTTAC-----GTCCAGGATACCCCTTAATGTA 245598
 Qy 2670 AGCATTATTTAATTAAGTTACTTGAATTAATGCTTATGCTGTATATTC-TGTTCA 2728
 Db 245597 AGCATTATTTTCTAGTTATTTGACT--TATTACATATTTATGTTGATACATTTCTTTTCA 245540
 Qy 2729 TCCATCGAATTTTCCCAAAAAGTAAAGATAGAGATAGGAGTACGCTTACATGCCAAGAAACT 2788
 Db 245539 TCCACGGATGTTCCCAAGAACTAGGAGTCTAGGAGTACGCTTCCACTTGCCTCAAGAAACT 245480
 Qy 2789 ATAAATTTTACTCTTAAATTTTACTTCTGAGCAGCTTGTGTTTATCAAGTCTTTTTTG 2848
 Db 245479 ATAAATTTTACTCTTAAATTTTACTTCTGAGCAGCTTGTGTTTATCAAGTCTTTTTG 245421
 Qy 2849 AAGAGACAGCCTCTGAAATTTCTTCAATCTGATACAGTGTACCTTGTATTTAAACATTT 2908
 Db 245420 AAGNAGATTACCTGTGAAATTTCTTCAATCTGATACAGTGTACCTTGTATTTGACATTT 245361
 Qy 2909 GTAATGTTGTTTCAAGTTTATCATCTCTTTTCAATCTTTTATAGCAAAATCAAAAGTATTAGC 2968
 Db 245360 GTAA---CGTGTCAAGTTTGTCTTCTTCTTCTTCTTCTTATAGCAATCAAACTTGTAGC 245304
 Qy 2969 TTCAGAAATTTATCAGAAATTTATATAAATTTTTCAGAGGTAAAGGCTTTTTTG 3028
 Db 245303 TTTAGAAATATATCAGAAATTTATATAAATTTTTCAGAGGTAAAGGCTTTTTTG 245244
 Qy 3029 TTAATAAATAAATAA---ATTATTTATTTTCTCTGATGAATAGAGGCTTTTTTATGCTG 3084
 Db 245243 TTAATAAATAAATAAATGATTTATTTATACCTTTCTTCTGTTGATAGAGGTTCTTTTATGCTG 245184

Qy 3085 CTGCTAATGAACCTAATTAGCTTAAATTTCTCTAGCAACATTTGGTCACTTTCAATC 3144
 Db 245183 CTGCTAATGAACCTAATTAGCTTAAATTTCTCTAGCAACATTTGGTCACTTTCAATC 245124
 Qy 3145 ATGCTATTAGCAAAAAA 316
 Db 245123 CCCTATTGCAAGAA 24506
 RESULT 8
 AJ699423
 LOCUS Rattus norvegicus mRNA for alpha-2,8-sialyltransferase (SIAT8F
 DEFINITION Rattus norvegicus mRNA for alpha-2,8-sialyltransferase (SIAT8F
 gene).
 ACCESSION AJ699423
 VERSION AJ699423.1 GI:47057304
 KEYWORDS alpha-2,8-sialyltransferase; ST8Sia-VI.
 SOURCE Rattus norvegicus (Norway rat)
 ORGANISM Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
 Sciurognathi; Muridae; Rattus.
 REFERENCE 1
 AUTHORS Harduin-Lepers, A., Molliou, R., Delannoy, P. and Oriol, R.
 TITLE The animal sialyltransferases and sialyltransferase-related genes:
 a phylogenetic approach
 JOURNAL Glycobiology 15 (8), 805-817 (2005)
 PUBMED 15843597
 REFERENCE 2 (bases 1 to 1197)
 AUTHORS Oriol, R.
 TITLE Direct Submission
 JOURNAL Submitted (04-MAY-2004) Oriol, R., US04, Inserm, 16 Av. Paul
 Vallant-Couturier, 94800, FRANCE
 FEATURES
 source
 1. .1197
 /organism="Rattus norvegicus"
 /mol_type="mRNA"
 /strain="Sprague-Dawley"
 /db_xref="taxon:10116"
 /clone="ST8Sia-1"
 1. .1197
 /gene="SIAT8P"
 1. .1197
 /gene="SIAT8P"
 /function="adds sialic acid to another sialic acid"
 /codon_start=1
 /product="alpha-2,8-sialyltransferase"
 /protein_id="CA27885.1"
 /db_xref="GI:47057305"
 /db_xref="GOA:Q2XC7"
 /db_xref="InterPro:IPR01675"
 /db_xref="InterPro:IPR012163"
 /db_xref="UniProtKB/TrEMBL:Q6ZXC7"
 /translation="MPPGTLFALVGLSMLLLLRFLWCPADAPARSMLGEESEED
 GFTSAALKTLSPPTVPTSTNLYDEKTPETAEKCKGLKYLSDSLANKTTRYSEDD
 YLQIVTNIQPCPKRLEEYENFRAKLASCDAIQDFVSNQNTPVGSNTTYVESKK
 QIPREINRMPFVSVNPSITLKYKLNKKKARFLIEDISAYGDAFLLLPFAFVYN
 TGSASQDVGSKTNLVNPSITLKYKLNKKKARFLIEDISAYGDAFLLLPFAFVYN
 TGSIFPKYQTLKESKRQKVLFPYRLHLALFWRTGVTAYRLSLGLMISIAVEL
 CENVTLGFWPFPSKTEEIPLSHHYYDNKLPKHGFHQPXKEYSQMLQLHMKGLKQF
 SKCAA"
 ORIGIN
 Query Match 32.5%; Score 1028; DB 6; Length 1197;
 Best Local Similarity 91.2%; Pos. No. 0;
 Matches 1091, Conservative 0; Mismatches 105; Indels 0; Gaps 0;
 Qy 77 ATGAGATCGGGGGGACGCTGTTCGCCCTCATAGGAGCCTGATGCTGCTCTCTCTG 136
 Db 1 ATGCGACCCCGGGGTACGCTGTTCGCCCTCATAGGAGCCTGATGCTGCTCTCTG 60
 Qy 137 CGTATGCTCTGCTGCCCGCAGCGGTCGCCCTCGCCGCTCCAGGCTGTTGATGAGGAGGAAGC 196

Db	1141	CAGCTCCATATGAAGGATCTT	CAAACTGAGTTACGAAATGTGAAGCGGCTTA	1196
Db	1141	CAGCTCCATATGAAGGATCTT	CAAACTGAGTTACGAAATGTGAAGCGGCTTA	1196
RESULT 9				
AJ621583				
LOCUS				
DEFINITION				
ACCESSION				
VERSION				
KEYWORDS				
SOURCE				
ORGANISM				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
FEATURES				
source				
gene				
CDS				
ORIGIN				
Query Match	27.28	Score	861.8	DB 5; Length 1234;
Best Local Similarity	82.08	Pos. No.	3.9e-271	
Matches 1006	Conservative	0	Mismatches 217	Indels 4; Gaps 1;
QY	77	ATGAGATCGGGGGGACGCTGT	CGCCCTCATAGGACGCTGTATGCTGCTGCTCTCTCTCTG	136
Db	1	ATGCGCGGGGGGCGACCTGT	CGCCCTGCTCGCAGGCTGCTGCTGCTGCTGCTGCTG	60
QY	137	CGTATGCTCTGTCGCCAGCCG	CGCGCTGCGCCGTCACAGGCTGTGTATGAGGGAAGC	196
Db	61	CGCTGCTCTGTCGCCAGCCG	CGCGCCGCGCGCGCCAGGATTCGTGAGGGAAGC	120
QY	197	AGAGAGACACCACTGTGTACCT	AGCTGCACTGAAGACACTCTGGAGCCGACCAACCCG	256

121 AGGAGGACCCACCGGACCCCGGAGCGCTCGGAGCGCCGCGACCGG 180
 257 GTACACGACGAGGACGACGACATATCTGGATGAGAGACACCCCAATAACAGAGAAA 316
 181 GTACCGCGCGGACCTAACAGACACATATCTGAATGAGAGTCTGCTCAACTGACGAGAAA 240
 317 TGCRAAGATCTGCAATATAGCTTGAACCTTTATCTAACAAAACGAGAGCTCTGAG 376
 241 TGCAAAATCTGCAATATGCGATTGAGTCTTTCTTAACAAAACGAGAGGTTATTCAGAG 300
 377 GATGACTACCTCCAGACCATCAACAAACATACAGAGATGCCATGCAACCGGCAACGAGAA 436
 301 AACGACTACCTTCAGATTATACAGATATACAGAGTTGTCATGAGAAACGAGCAAGAGAA 360
 437 GAATATGACAAATTTAGAGCAAACTGGCTTCTGTTGCGATGCGCATTCAGACATTCGTG 496
 361 GAATATGCAAAATTTAGAGCCAACTTGTCTTCTGCTGTGATGCTGTTCAAAATCTTGT 420
 497 GTTTCCCGAGAACACACTCCAGTGGGACTACATGAGTACGAGTGGAGTGAAGCAAGAA 556
 421 GTTTCTCAGAAATACACTCCAGTGGGACTAATATGAGTTAGAGTGGAGTGAAGCAAGAA 480
 557 CACATCCCATTCGAGAGAACATTTTCCACATGTTTCCAGTGTGCGAGCTTTTGTGAC 616
 481 GAATCCCAATTAAGAGAACATTTTATATGTTTCCAGTGTGCGAGCTTTTGTGAC 540
 617 TATCCCTATAACAGTGTGAGTGGTGAATGGGGAATTTCTCAACAGTCTCTCTGCG 676
 541 TACCTTTAATACAGTGTGAGTGGTGAATGGGGAATTTCTGAATAGTCTCTCTGT 600
 677 GGAGCAGAAATGTAAATCTGACTTCTGCTTCCAGTGTGACCTCCCGGCAATCACAGG 736
 601 GGAATGAAATAGATAAATTCGACTTCTGTTTGGTGTAACTTACCTCCCGGCAATCACAGG 660
 737 AGCGTACTGAAAGTGTGGAGCAAAACAAATCTTGTGACTGTCAATCCCGGCAATTA 796
 661 GATGTTAGTAAAGATTTGGCAGTAAACAAATCTTGTGACTATAATCCAGCATCAT 720
 797 ACCCTGAAGTACCGAATTTGAAGGAGAAAGACACAGTGTGAGGAGCATCTCCACC 856
 721 ACTCTGAATATGGGAATTTAAAGGAAAAAAGCCCTTCTTCTGAGGAGCATTCGAACC 780
 857 TATGAGAGTACATCTCTCTGCGACATTTTCTATCGGCGCAACACAGGCAATCTCT 916
 781 TATGAGAGTACATTTTCTCTGCGACATTTTCTTCAAGGCGCAACACAGGCAATCTCT 840
 917 TTTAAAGTCTACCAAACTCAAGAGTCAAAATGAGGCAAAAGGTTCTCTCTTCCAT 976
 841 TTCAAAGTATATACAGCTCGAAGGCTTAAGCAAGCAAAAGGTTCTATTTTCCAT 900
 977 CCCAGGTACCTGAGACACCTCGCTCTTCTGAGAACTAAAGGGGTGACTGCATACCGC 1036
 901 CCCAAGTACCTGAAAGATCTGGCCCTTTTCTGGAGAACTAAAGGTGTGACTGCATACCGC 960
 1037 TTGTCCACAGCTTGATGCAATGCAAGTCTGCTGGAAGTGTGGAAGGTAAGGTAAGTGC 1096
 961 TTGTCCACCGCTTGATGATCAAGTGTGCAAGTGTGCAAGTGTGTAAGGTAAGTGTG 1020
 1097 TACGATTTCTGCGCTTTCTCTAAAGACTATCGAAGACACCCCACTCAGTCAACACTACT 1156
 1021 TATGATTTCTGCGCTTTCTCTAAAGCTGTAGAGACATACCTGTGAGCCATCACTATTAT 1080
 1157 GATAACATGTTACCTAAGCATGTTTCCACCAAGATGCTTAAAGAAATACAGGCAAAATGTC 1216
 1081 GACAACAGCTACCTAAACATGTTTCCATCAGATGTCGCAAGAAATACAGGCAATCTCT 1140
 1217 CAGCTCCATATGAGAGGAATCTCTAACTGCAATTCAGCAAAATGTGAACCGGCTT---A 1272
 1141 CAACTTCAATGAAGAGGAATCTCTAACTGCAATTTAGCAAAATGTGAACCGGCTTAAACA 1200
 1273 ACGTTTCTTAGAGGAGGAATTTTCA 1299
 1201 AAGTATCTTAAATGGGAATTAATTTTA 1227

RESULT 10
 AJ697663
 LOCUS
 DEFINITION Pan troglodytes mRNA for alpha-2,8-sialyltransferase 8P (sial8P gene)
 ACCESSION AJ697663
 VERSION alpha-2,8-sialyltransferase 8P; sial8P gene.
 KEYWORDS Pan troglodytes (chimpanzee)
 SOURCE Pan troglodytes
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominoidea; Pan.
 REFERENCE 1
 AUTHORS Harduin-Lepers, A., Mollion, R., Delannoy, P. and Oriol, R.
 TITLE The animal sialyltransferases and sialyltransferase-related genes: a phylogenetic approach
 JOURNAL Glycobiology 15 (8), 805-817 (2005)
 PUBMED 15843597
 REFERENCE 2 (bases 1 to 1197)
 AUTHORS Oriol, R.
 TITLE Direct Submission
 JOURNAL Submitted (20-APR-2004)
 FEATURES
 source
 1. .1197
 /organism="Pan troglodytes"
 /mol_type="mRNA"
 /db_xref="taxon:9598"
 1. .1197
 /gene="sial8P"
 1. .1197
 /gene="sial8P"
 /function="adds sialic acid to another sialic acid"
 /codon_start=1
 /product="alpha-2,8-sialyltransferase 8P"
 /protein_id="CA26901.1"
 /db_xref="GI:46117917"
 /db_xref="GOA:P1648"
 /db_xref="UniProtKB/Swiss-Prot:P61648"
 /translation="MPEGALLALLASILLILLILLWCPADAPGRARILVEESREATHPTAALTRSPATVPRATNSYLNEKSLHLEKCNLQYIGIESFNKTYGSENDYLIQITDIQCPMKRABEYANFRKLASCCDVQNFVVSQNNPTVGNMTSEVSKK EIP1KNI FHMFPVS PFVDYVYQCAVVGNGGILNKLSCGTEIDKSDVFRCPNLPPT TGVSKDVGKSTNLV INPSIITLYGNLKEKALFLIEDIATYGEAPFLPAPFRAN TGTSFKVYVTLERSK ROKVLFPHPKYLDLALFWRTKGTAYRLSTGLMTTSVAVEL CKNKVLYGFVWPPSKT EDIPVSHHYDNLKLPKHGHPMPKPSYQILQLHMKGLKLP SKCEVA"
 ORIGIN
 Query Match 27.1%; Score 857.8; DB 5; Length 1197;
 Best Local Similarity 82.3%; Pr. Id. No. 8e-270;
 Matches 985; Conservative 0; Mismatches 212; Indels 0; Gaps 0;
 QY 77 ATGAGATCGGGGGGCGACGCTGTTCGCCCTCATAGCAGCCTGATGCTGCTCTCTCTG 136
 Db 1 ATCGCGCGCGGGGGCGGCGACTGTCGCCCTGCTCGCCAGCCTGCTGCTGCTGCTG 60
 QY 137 CGTATGCTCTGCTGCCCGGCGGCGCGCGCTGCTCGAGCTGTTGATGAGGAGGAGC 196
 Db 61 CGCTGCTCTGCTGCCCGGCGGCGCGCGCTGCTCGAGCTGTTGATGAGGAGGAGC 120
 QY 197 AGAGAGACACACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 256
 Db 121 AGGAGAGCACCACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 180
 QY 257 GTACCAACGACCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 316
 Db 181 GTACCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 240
 QY 317 TCGAAGATCTGCAATATAGCTTAACTTTTATCTAACAAAACGAGAGCGGTACTCTG 376

Db	361	GAATATGAGATTTT	AGAGCAAAATCTCGCTCTCTGTTGATGCTGTTTCAAACTTCATT	420
Qy	497	GTITCCGAGAACAA	CACATCTCAGTGGGACTAATCATGAGCTACGAGTGGAAAGCAAGAAA	556
Db	421	GTITCTCAAAACAA	CACATCTCCATTTGGACCAACATGATGATGAGTGGAAAGCAAGAAA	480
Qy	557	CACATCTCCATTT	CGAGAGAAATTTTCCATGATGTTTCCAGTGTGCGAGCTTTTCTGGAC	616
Db	481	AAATTCAAATTA	AGAAACAAATTTTGTATGTTTCCAGTGTGCGAGCTTTTCTGGAC	540
Qy	617	TATCCCTATAAC	CCAGTGTGAGTGTGTTGTAATGAGGGAATTTCTCAACAGTCTCTCTGC	676
Db	541	TACTCTTTTAA	TAACTGAGTGTGTTGTAATGAGGGAATTTTAAATCAGTCTCTCTGT	600
Qy	677	GGAGCAGAAATTA	GAATTAATCTGCTTTCAGTGTGTAACCTCCCGCAATCACAGGG	736
Db	601	GGAGCTGAATTC	GAATTAATCTGCTTTCAGTGTGTAACCTCCCGCAATCACAGGA	660
Qy	737	AGCGCTAGTAA	GATGTTTGGAGCAAAACAAATCTTGTGATGTTCAATCCCGCAATTA	796
Db	661	AATGTTACCA	TGATGTTTGGAGCAAAACAAATCTTGTGATGTTCAATCCCGCAATTA	720
Qy	797	ACCTGGAAGTAC	CAAGATTTTGAAGGAGAAAGACACAGTTTTTCGAGGACATCTCCACC	856
Db	721	AAGCTAAGAT	TGGAACTTAAAGGAAAGAAAGCAATTTTTCGAGGAGCTTCGACC	780
Qy	857	TATGAGATGCA	ATTTCTCTCTGTCAGATTTTCTATCGGCGCAACACAGGCAATCTCT	916
Db	781	TATGAGATGCT	TTTCTCTCTGTCAGATTTTCTATCGGCGCAACACAGGCAATCTCT	840
Qy	917	TTTAAAGTCTAC	CAAACTCAAGAGTCAAAATGAGGCAAAAGGTTCTCTCTTCTCCAT	976
Db	841	TTTAAAGTCTAC	CAAACTCAAGAGTCAAAATGAGGCAAAAGGTTCTCTCTTCTCCAT	900
Qy	977	CCCAGGTACCT	GAGACACCTCTCTTTCTGGAGAACTTAAAGGGGTGACTGCATACCGC	1036
Db	901	CCCAGGTACCT	GAGAAACCTTGCATCTTCTGGAGAACTTAAAGGGGTGACTGCATACCGC	960
Qy	1037	TGTCCACAGGCT	TGATGATGCAAGTGTGCTGTGGAACCTGTTGTAAGGTTGAGGATGCTC	1096
Db	961	CTGTCTCTCGG	CTTATGATGCAAGTGTGCTGTGGAACCTGTTGTAAGGTTGAGGATGCTC	1020
Qy	1097	TACGGATTCGG	CTTCTCTAAGACTATGAGACACCCCACTCAGTCAACCTACTACT	1156
Db	1021	TATGGATTCGG	CTTCTCTAAGACTATGAGACACCCCACTCAGTCAACCTACTACT	1080
Qy	1157	GATAACATGTT	TAACATGATGTTTCCACAGATGCTTAAAGGATGATGATGATGATGATGAT	1216
Db	1081	GACATAGTTAC	TAACATGATGTTTCCACAGATGCTTAAAGGATGATGATGATGATGATGAT	1140
Qy	1217	CAGCTCCATAT	GAGAGGAATCTCAAACTGCAATTCAGAAATGTAAGGATGTAAGGATGTA	1273
Db	1141	CAGCTTCATGT	GAAAGGATCTCTCAAGTTACAGTTTAGCAATGTAAGGATGTAAGGATGTA	1197
RESULT 12				
CQ736679				
LOCUS				
DEFINITION				
ACCESSION				
VERSION				
KEYWORDS				
SOURCE				
ORGANISM				
Homo sapiens (human)				
Homo sapiens				
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;				
Hominidae; Homo.				
1				
Venter, C.J., Adams, M.C., Li, P.W. and Myers, E.W.				
Kits, such as nucleic acid arrays, comprising a majority of				
human exons or transcripts, for detecting expression and other uses				
thereof				
Patent: WO 02068579-A 22613 06-SEP-2002;				

FEATURES		PE Corporation (NY) (US)	Location/Qualifiers	
source		1. 1149		
		/organism="Homo sapiens"		
		/mol_type="unasigned DNA"		
		/db_xref="taxon:9606"		
ORIGIN				
Query Match		17.7%; Score 559.8; DB 2; Length 1149;		
Best Local Similarity		75.5%; Pr.d. No. 1.1e-171;		
Matches 754; Conservative		0; Mismatches 152; Indels 93; Gaps 1;		
Qy	275	AGCACATATCTGGATGAGAGAG	AACCCAAATTAACAGAGAAATGCAAGATCTGCAATAT	334
Db	244	AGCACATATCTGGATGAGAGAG	GCTCCACTGACGGAGAAATGCAAAATCTGCAATAT	303
Qy	335	AGCTTGAACTCTTTATCTTAACA	AACGAGACGGTACTCTGAGGATGACTACCTCCAGACC	394
Db	304	GGCATTTGATCTTTCTCTAACA	AACGAAAGGGTATTTCAGAGAACGACTACCTTCAGATT	363
Qy	395	ATCAAAACATACAGAGATGCC	ATGGAACCGGCAAGCAGAGAAATATGACAAATTTTGA	454
Db	364	ATCAAGATATACAGAGTTGTC	ATGGAACCGGCAAGCAGAGAAATATGCAAAATTTTGA	423
Qy	455	GCBAACCTGGCTTCTCTGTGCG	TGCCAATTCAGACTTCTGTTGTTTCCAGAACCAACT	514
Db	424	GCCAACTTGTCTCTCTGTGTC	TGCTGTTCAAACTTTGTTTCTCAGAAATTAACACT	483
Qy	515	CCAGTGGGAGCTAAACATGAGCT	CGAGTGGAAAGCAAGAAACACATCCCCATTCGAGAG	574
Db	484	CCAGTGGGAGCTAATATGAGTT	CGAGTGGAAAGCAAGAAATCCCAATTAAGAAG	543
Qy	575	AACATTTTCCATGTTTCCAG	GTGCGAGCTTTTGTGGACTATCCCTATTAACAGATT	634
Db	544	AACATTTTCCATGTTTCCAG	GTGCGAGCTTTTGTGGACTATCCCTATTAACAGATT	603
Qy	635	GCAGTGTGTAATGAGGAGAA	TCTCAACAAAGTCTCTCGGAGCAGAAATTCATAAA	694
Db	604	GCAGTGTGTAATGAGGAGAA	TCTCAACAAAGTCTCTCGGAGCAGAAATTCATAAA	663
Qy	695	TCTGACTTCTGTTTCCAGTGT	CTCCGCCCAATTCACAGGGAGCGCTAGTAAAGATT	754
Db	664	TCCGACTTCTGTTTCCAG	CTCCGCCCAATTCACAGGGAGCGCTAGTAAAGATT	680
Qy	755	GGAAGCAAAACAAATCTTTGTA	TGTCATCCAGCATTAATAACCTGAAGTACAGAAAT	814
Db	681	-----	-----ATATGGGAAC	690
Qy	815	TTGAGGAGAGAGAGAGCAGAT	TTTGGAGGACATCTCCACTATGGAGATGCAATTCCTC	874
Db	691	TTAAGGAGAGAGAGAGCAGAT	TTTGGAGGACATCTCCACTATGGAGATGCAATTCCTC	750
Qy	875	CTCCTGCGCAGATTTTCTCTATC	GGCCAAACACAGGCACTCTCTTTTAAAGTCTACCAACA	934
Db	751	CTTCTGCGCAGATTTTCTCTTCA	GGCCAAACACAGGCACTCTCTTTCAAGTATCTACACG	810
Qy	935	CTCAAGAGTCAAAATATGAGGC	AAAGTTCTCTTTTCCATCCAGTACCTGAGACAC	994
Db	811	CTCAAGAGTCAAAATATGAGGC	AAAGTTCTCTTTTCCATCCAGTACCTGAGACAC	870
Qy	995	CTCGCTCTTTTCTGAGAACTA	AGGGGTGACTGCATACCGCTGTTCCACAGGCTTGATG	1054
Db	871	CTCGCTCTTTTCTGAGAACTA	AGGGGTGACTGCATACCGCTGTTCCACAGGCTTGATG	930
Qy	1055	ATTCCAAGTGTGCTGTGGAACT	GTGTGAAACCGTGAAGCTCTACGGATTTCTGGCTTTC	1114
Db	931	ATCAAGAGTGTGAGTGGAACT	GTGTGAAACCGTGAAGCTCTACGGATTTCTGGCTTTC	990
Qy	1115	TCTAAGACTATCGAAGACACCC	ACTCAGTCCACACTACTATGATAAATGATTAACCTAAG	1174
Db	991	TCTAAGACTATCGAAGACACAC	ACTCAGTCCACACTACTATGATAAATGATTAACCTAAG	1050
Qy	1175	CATGGTTTCCACAGATGCCTTA	AGAATACAGCCAAATGCTCCAGCTCCATATGAGAGA	1234

```

|||||
1051 CATGTTTCCATCATGTCCTCCCAAGATACACCGAGATCCTCAACTTCACATGAAGGA 1110
|||||
1235 ATCCTCAAACTGCAATTCAGCAAAATGTGAAACGGCTTAA 1273
|||||
1111 ATCCTCAAACTGCAATTTAGCAAAATGTGAAGTCGCTTAA 1149
|||||

```

RESULT 13
AL158164/c
LOCUS
DEFINITION Human DNA sequence from clone RP11-414K1 on chromosome 10 Contains the 3' end of the gene for the likely ortholog of mouse sialyltransferase 8-VI (alpha-2 8-sialyltransferase) (ST8SIA-VI), a novel gene and the 5' end of a novel gene, complete sequence.

ACCESSION AL158164
VERSION AL158164.15 GI:16416157
KEYWORDS HTG; sialyltransferase; ST8SIA-VI.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 151289)
Pearce, A.

Direct Submission
Submitted (13-MAY-2005) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: vegasanger.ac.uk
Clone requests: clonerequest@sanger.ac.uk
On Oct 24, 2001 this sequence version replaced gi:15028618.
The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em, EMBL; Sw, SWISSPROT; Tr, TrEMBL; Wp, WORMPEP; Information on the WORMPEP database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome 10, constructed by the Sanger Centre Chromosome 10 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/Chrio>
RP11-414K1 is from the library RPEC11-2 constructed by the group of Pieter de Jong. For further details see <http://www.chori.org/bacpac/home.htm>
VECTOR: pBACe3.6

----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: <http://www.sanger.ac.uk>
Contact: vegas@sanger.ac.uk

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one subclone; and the assembly was confirmed by restriction digest, except on the rare occasion of the clone being a YAC.

FEATURES
source
1. 151289
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="10"
/clone="RP11-414K1"
/clone_lib="RPEC11-2"
/clone="RP11-414K1"
2000
/note="Clone right end: RP11-124N14"
complement(join(<54812. 55162,57619. 57775,59494. 59586, 63443. >63538))
/gene="RP11-414K1.1"
/locus_tag="RP11-414K1.1-002"
complement(join(<54812. 55162,57619. 57775,59494. 59586, 63443. >63538))
/gene="RP11-414K1.1"

misc_feature
/note="Clone right end: RP11-124N14"
complement(join(91406. 91512,91566. 91792,95943. 96029, 103847. 103960, 126960. 126995))
/gene="RP11-414K1.1"
/locus_tag="RP11-414K1.1-003"
complement(join(91406. 91512,91566. 91792,95943. 96029, 103847. 103960, 126960. 126995))
/gene="RP11-414K1.1"

gene
/locus_tag="RP11-414K1.1-002"
complement(join(91406. 91512,91566. 91792,95943. 96029, 103847. 103960, 126960. 126995))
/gene="RP11-414K1.1"

mRNA
/locus_tag="RP11-414K1.1-003"
ortholog of mouse sialyltransferase 8-VI

/locus_tag="RP11-414K1.1-002"
ortholog of mouse sialyltransferase 8-VI
/product="likely ortholog of mouse sialyltransferase 8-VI (alpha-2, 8-sialyltransferase) (ST8SIA-VI)"
/note="match: E:PF678643.1"
complement(join(55124. 55162,57619. 57775,59494. 59586, 63443. >63538))
/gene="RP11-414K1.1-002"
/locus_tag="RP11-414K1.1-002"
/standard_name="OTTHUMP00000019236"
/codon_start=2
/product="likely ortholog of mouse sialyltransferase 8-VI (alpha-2, 8-sialyltransferase) (ST8SIA-VI)"
/protein_id="CA40659.1"
/db_xref="GI:5708950"
/db_xref="GOA:Q0JUL6"
/db_xref="UniProtKB/TrEMBL:Q5JUL6"
/translation="YFNQCAVGVNGGILNKSLCCTGTEIDKSDVFRCNLPPTTGDVSK DVGSKTLVLTINPSI ILTKGNLKEKALFLEDIATYGDAPFLPAPFRANTGTSPK VYTLSESKARQCKC KRKTINSIL"
join(complement(1160289. 14:41295. 41470), complement(1160289. 14:40699. 40797), complement(12690. 127049), complement(95943. 96029), complement(6783. 67981), complement(63443. 63555), complement(5949. 59586), complement(56303. 57775))
/gene="RP11-414K1.1-001"
/locus_tag="RP11-414K1.1-001"
join(complement(1160289. 14:41295. 41470), complement(1160289. 14:40699. 40797), complement(12690. 127049), complement(95943. 96029), complement(6783. 67981), complement(63443. 63555), complement(5949. 59586), complement(56303. 57775))
/gene="RP11-414K1.1-001"
/locus_tag="RP11-414K1.1-001"
ortholog of mouse sialyltransferase 8-VI
/product="likely ortholog of mouse sialyltransferase 8-VI (alpha-2, 8-sialyltransferase) (ST8SIA-VI)"
/note="match: E:PF677781.1 BB660248.1 BS686184.1 BG19404.1 B110002.1 match: CDNAS: A:059554.1"
join(complement(1160289. 14:41295. 41395), complement(1160289. 14:40699. 40797), complement(12690. 127049), complement(95943. 96029), complement(6783. 67981), complement(63443. 63555), complement(5949. 59586), complement(57307. 57775))
/gene="RP11-414K1.1-001"
/locus_tag="RP11-414K1.1-001"
/standard_name="OTTHUMP00000019235"
/note="match: P:PF0554 P79783 P7877 Q64468 Q64687 Q87ZQ3 Q8K4T1 Q92185 Q9CUJ6"
/codon_start=1
/product="likely ortholog of mouse sialyltransferase 8-VI (alpha-2, 8-sialyltransferase) (ST8SIA-VI)"
/protein_id="CA40658.1"
/db_xref="GI:5708949"
/translation="M PGGALLIALSILLILLRLLWCPADAPGARILVEESREAT HPTPALRTLSRSPAT VPRATNLYLNKSLQLECKNLOVIGTESFNKTKGYSEND YLQITDIQCPWKR ABEYANFRAKLASCCDAVQNVFVSNPTFGVTNMYEVESKK EIPKKNIFHFPPVS PFVDYPYQCAVGVNGGILNKSLCCTGTEIDKSDVFRCNLPPT TGDVSKDVGSKTLNVLNPSI ILTKGNLKEKALFLEDIATYGDAPFLPAPFRANTGTSPK VYTLSESKARQCKC KRKTINSIL RQKVLFFHPKYLKDIALFWRTKGTAVRLSLTIMITSVAVEL CKVKLYGVFWPFSKTD EDIPVSHHYDNKLPKHGFHQMPKYSQILQLHMKILQKF SKCEVA"
85117
/note="Clone left end: RP11-161K20"
complement(join(91406. 91512,91566. 91792,95943. 96029, 103847. 103960, 126960. 126995))
/gene="RP11-414K1.1"
/locus_tag="RP11-414K1.1-003"
complement(join(91406. 91512,91566. 91792,95943. 96029, 103847. 103960, 126960. 126995))
/gene="RP11-414K1.1"

misc_feature

gene

mRNA

```
(alpha-2, 8-sialyltransferase) (ST8SIA-VI)
/nc=match: ESTs: AV714822.1"
Join(123423..123504,135534..135608,144397..144715)
/locus_tag="RP11-414K1.3-001"
Join(123423..123504,135534..135608,144492..144715)
/locus_tag="RP11-414K1.3-002"
Join(123423..123504,135534..135608,144397..144715)
/locus_tag="RP11-414K1.3-001"
/product="novel transcript"
/nc=match: ESTs: AL953053.1 AW139127.1 AW340196.1
B1826833.1 BU537392.1
match: cDNAs: BC029655.1"
Join(123423..123504,135534..135608,144492..144715)
/locus_tag="RP11-414K1.3-002"
/product="novel transcript"
/nc=match: ESTs: AA813705.1 AI6322435.1 AM449205.1
BE045987.1"
Join(144469..144673,149779..149932)
/locus_tag="RP11-414K1.2-001"
Join(144469..144673,149779..149932)
/locus_tag="RP11-414K1.2-001"
/product="putative novel transcript"
/nc=match: ESTs: AA609183.1"
144692..144697
/locus_tag="RP11-414K1.2-001"
144715
/locus_tag="RP11-414K1.2-001"
151289
/nc="Clone_right_end: RP11-414K1"

ORIGIN
Query Match 14.9%; Score 470.4; DB 5; Length 151289;
Best Local Similarity 64.7%; Pred. No. 6.1e-142; Indels 69; Gaps 18;
Matches 1008; Conservative 0; Mismatches 481;

Qy 740 GCTAGTAAAGATGTTGGAGCAAAACAAATCTGTGACTGTCAATCCCGACGATTATAACC 799
Db 57840 GTTATCAAACTGCTACTTCTGATTAATTAATCTTGGATTACTATAATTTTTTTTAAATT 57781

Qy 800 CTGAAGTACCAAGATTGAAGAGAGAAAGACAGATTTTGGAGGACATCTCCACCTAT 859
Db 57780 GGCAGATATGGAACTTAAAGGAAAAAAGCCCTATTCTCGAGGACATTGCAACCTAT 57721

Qy 860 GGAGATGATTCCTCCTCTGCCAGCATTTTCCATCGGGCCCAACACAGGATCTCTTTT 919
Db 57720 GGAGATGATTTTCTTCTGCCAGCATTTTCTCGAGGCACTAAAGGTTCTCTTCCATCCC 979

Qy 920 AAAGTCTACCAAACTCAAAGAGTCAAAATGAGGCAAAAGTTCTCTTCTTCCATCCC 979
Db 57660 AAAGTATACTACAGCTCGAAGAGTCTAAAGCAAGCAAAAGTTCTATTTTCCATCCC 57601

Qy 980 AGGTACCTGAGACACCTCGCTCTTTTCTGGAGAACTAAAGGGGTGACTGCATACCGCTG 1039
Db 57600 AAGTACCTGAAAGATCTGGCCCTTTTCTGGAGAACTAAAGGTTGACTGCATACCGCTG 57541

Qy 1040 TCCACAGCTTGATGATGCAAGTGTCCGTGTGAACTGTGTGAAAGCTGAAGCTCTAC 1099
Db 57540 TCCACAGCTTGATGATGCAAGTGTGCAAGTGTGCAAGTGTGTAAGAAATGTAAGCTGTAT 57481

Qy 1100 GGATTTGGCTTTCTCTTAAGACTATCAAGACACCCCACTCAGTCCACCTACTATGAT 1159
Db 57480 GGATTTGGCTTTCTCTTAAGACTATCAAGACACCTCAGTCCACCTACTATGAT 57421

Qy 1160 AACATGTTACCTAAGCAGTGTTCACCAAGATGCTTAAAGAAATACAGCAAAATGCTCCAG 1219
Db 57420 AACAGCTACCTAAGCAGTGTTCACCAAGATGCTTAAAGAAATACAGCAAAATGCTCCAA 57361

Qy 1220 CTCATATGAGAGGAATCTCAAACTGCAATTCAGCAAAATGTGAAACGGCTT----AAGC 1275
Db 57360 CTCATATGAGAGGAATCTCAAACTGCAATTCAGCAAAATGTGAAACGGCTT----AAGC 57301

Qy 1276 TTTCTTAGAAGAGAAATTTTCAGGA-GGTGGAGTGGATGTGTCACAGCATCTCCAAA 1334
Db 1334 TTTCTTAGAAGAGAAATTTTCAGGA-GGTGGAGTGGATGTGTCACAGCATCTCCAAA 1334
```

RESULT 14
AJ699424
LOCUS
DEFINITION
Gallus gallus mRNA for alpha-2,8-sialyltransferase (SIAT8F gene).
ACCSSION
AJ699424
VERSION
AJ699424.1. GI:47057306

```
Db 57300 TATCTTAAATGGGAATAATTTTAAATAATGACAGTAGTGATTAACAATGCTCTCCAAAC 57241
Qy 1335 AGCCATATAGAGAGGCGACAGAATAAGCATGAATTAAGAGGCGTCTCTCCACTGTCTTA 1394
Db 57240 ACC-----AAAGGAGGTGGTAAAGAGATATTTTGGATGAGGCCCAAAATTTGGTTT 57189
Qy 1395 GACCAAGCCACCGCCGCCACCACTTTGAGCCTCCACAGAGTCACTC-ATTCTCACCT 1453
Db 57188 GACCAAG-----CTTCCCCACCAATTTGCAATGATGGCAAGTCAATCAATCTCTCTCA 57134
Qy 1454 TCAAGCTTTCTTCTCTGAGAATAGACAAAACATCAGACTTGGATAGTAAATGAGA 1513
Db 57133 TCTTCATTTTCTCTCTTATTAATGAGACACCATATCTGACTCATACAAATTTTATAAT 57074
Qy 1514 TAATTTTCAATCATCATAGATTGATTTGAGCCAGGCTCTCTCAGATGCTCTCTTG 1573
Db 57073 TTATGAAAAATTAATGGCATGGCTTTTGGTGTGAGTAGGAGACTCAAGAATAATTTCTTTC 57014
Qy 1574 TTCCTATCCATGATAGCCATTTCCACCTTTATCAGAGTGGTAAGAACTGTGCAATTTCT 1633
Db 57013 TTACTATTTATACAGCATCTCACCTTCCAGCAGCAATTAATAAGCATGCTTTCTT 56954
Qy 1634 GCCAAAGACCTTTCTGAGAGATGCTCTGAAT-----CATGCCCGAGTTTTTACACA 1687
Db 56953 G---GAGATCTCTTCTTAAAGAGATAGCTGCATTACCAAGTATCCATTGCTTTCTG 56897
Qy 1688 CAGCTCTCTCTTTATAATAATAAAGCTTCCCATCTCTCTCTAGTAGAGTACAG----- 1741
Db 56896 TGGCAGGCTCTCTATCAATAATAAGCTTCCCATTTTCTCTCTCAAGAGCTGCATCAAGA 56837
Qy 1742 -----AAACAAAATCCCTTGATGATTACAGGAAGAAAAGTCTTTTTTACTTTA 1788
Db 56836 AAATATATATCTTAAAGCAAAAATCTTGATCATTTGGGAATGAATTTCTTCTCGAGTA 56777
Qy 1789 --GCAATGTGCTCTCTCTGATCAGTTGCTGTGACATTAAGCTGGGTGGGTTTGG 1846
Db 56776 TGGCAGATGCTGCTCTCTGATAAATTCATTTTAAAGT-AGCTGTGATGATATTTG 56718
Qy 1847 GTTGGATTTGGGGGTTTCTTCTCTTTTGTCTATATTTTCTTACCTTTATCAGTTT 1906
Db 56717 GTTGAATTTAAGCCATTTTATTTCTCTATGTCTACATTTCTTCATCACTTAATAATTT 56658
Qy 1907 GTATTTCAGAGCTTCTGCTTTTGGATTTGCAATTTCTCTCCACTGACAGATCAACTC 1966
Db 56657 -TATTTAACTTCTGCTCTCATCTTACACTTTTCTCTCTCTATACGATATACTT 56599
Qy 1967 AATGACATAAAGTAGTTTCAAACTCCCATTTGCTCTCACAATGTTTTAT---CCATAAAGTT 2023
Db 56598 CAAGGACGTGAACAGTTACAAATCCAGTCTTTCTTACCTGTGTGCTTGGCTTGGGTC 56539
Qy 2024 ACTCATCTGATTTTATTTAAAAATAGTGAAATCTACTTGATATACAGCCGAGGACCA 2081
Db 56538 ACTCACTCAATTTTATTTTAAATTTACTGAATAACTACTTCAATGCGAGGCCCT-GCACCA 56480
Qy 2082 TCTTCCATTTGAGAGATATGAG--TATTGTCACTGGCAGAAAAAGCAGGTGTGCCATT 2139
Db 56479 GGCTCTGTGAGAGGACAGAGATTATTTGCTCTGTAGAGAGAAACAGATGTGCTAT 56420
Qy 2140 AATTGATAA-----GATACCAAGCATCATATGCC---AGTTATGAACACAGTGTG 2190
Db 56419 CATTAATTTGATGGGTATGAAGTATTACAAATGACAAAAGCTATGAATAAAGTGCTA 56360
Qy 2191 AAGGATCATAGACGGGTGGTAAATCTGATCCAGTAGAATAAATCTGATGCTGAC 2248
Db 56359 CAAGGAACAGAGCAGGAGTAAATAATC-CATTTAGTGGAAATCTATTTCAATGTAC 56303
```

KEYWORDS	alpha-2,8-sialyltransferase; SIAT8F gene.
SOURCE	Gallus gallus (chicken)
ORGANISM	Gallus gallus
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.
AUTHORS	Harduin-Leper, A., Mollicon, R., Delannoy, P. and Oriol, R.
TITLE	The animal sialyltransferases and sialyltransferase-related genes: a phylogenetic approach
JOURNAL	Glycobiology 15 (8), 805-817 (2005)
PUBMED	15843597
REFERENCE	2 (bases 1 to 1197)
AUTHORS	Oriol, R.
TITLE	Direct Submission
JOURNAL	Submitted (04-MAY-2004) Oriol R., U504, Inserm, 16 Av. Paul Vaillant-Couturier, 94807, FRANCE
FEATURES	Location/Qualifiers
source	1..1197
gene	/organism="Gallus gallus"
CDS	/mol_type="mRNA"
	/db_xref="taxon:9031"
	/clone="ST8Sia-VI"
	1..1197
	/gene="SIAT8F"
	1..1197
	/gene="SIAT8F"
	/function="adds sialic acid to another sialic acid"
	/product="alpha-2,8-sialyltransferase"
	/protein_id="CAG27886.1"
	/db_xref="GI:47057307"
	/db_xref="GOA:O62XC6"
	/db_xref="InterPro:IPR011675"
	/db_xref="InterPro:IPR012163"
	/db_xref="UniProtKB/TrEMBL:O62XC6"
	/translation="MPAGQPSGSRPVAALAGDFNFCYPVLLARGWEEAEVATVT PKVVALRSPLTSLPOTENRYTANKDGIQFELDTKCAIQDNLSSFKKRPEDY YLHIVKIQNCTWVRPEESAKFRELASCAVGNVNFIAQONTPGLSGNMSYVENDSK TILITIEDFRLMPLSPVPLSPYFKNCVAVGNGILKNSCGAEIDRSDFVFRCLPPT MGSISKQVNTKLVTVNPSIIAOKYNKLEKTEFVENVAVYDAFLLLPAFPRSN TATSPKVHTLQEPKATORAIFPHPTYLKILNLAQFWRTKGVKAYRLSSGFMITSAAVEL CENVKLYGFWPSPSKSTKMPISHHYYDNLQPKPGHAMPKEYNQILQLHGKIKLQF GKCESD"
ORIGIN	
Query Match	14.8%; Score 467.2; DB 11; Length 1197;
Best Local Similarity	67.9%; Pred. No. 3.5e-141;
Matches 552; Conservative	0; Mismatches 308; Indels 0; Gaps 0;
Qy	314 AAATGCAAGATCTCAATATAGTTGAACCTTTTATCTTAAACAAACGAGACGGTACTCT 373
Db	238 AAATGTAAGGCAATTCAGGACCAACATTTTGTTCATGCTGCTGCTTCAAGAAAGGATATCCA 297
Qy	374 GAGGATGACTACTCCAGACCATCAACATACAGAGATGCCATGACCGGCAAGCA 433
Db	298 GAAGATTATATCTTCATACATAGTCAAAAACATGCGAAGTGCACCTGGGTAGGAGACCG 357
Qy	434 GAAGATATGACAAATTTAGACAAAACATGGCTTCTGTTGGCATGCCATTCAGACATTC 493
Db	358 GAAGATCAGCAAAATTCAGGTACAGTTAGTTCTGCTGTGTAATGCAGTTTCAATTTT 417
Qy	494 GTGGTTTCCGAAACAACTCCAGTGGGAGTAAACATGAGTACGAGTGGAAAGCAAG 553
Db	418 ATTGCTTCTCAAAACAAACACCCACTGGGAAGTAAATGAGTATGAAAGTGACAGTAAA 477
Qy	554 AAACATCCCATTCGAGAGACATTTTCCACATGTTCCAGTGTCCGAGCTTTTGTG 613
Db	478 AAGACCTCTCATACAGAGACATTTTAAAGATGTCGCTGTGCTCTCTCTCTTCA 537
Qy	614 GACTATCCCTAFAACAGTGTGAGTGGTGGTAAATGGGGAAATTCCTCAACAGTCTCTC 673
Db	538 GTTATCTCTTCAAGAACTGTGCACTGGTTGGAATGAGGCAATTCGAAAAATTCACG 597
Qy	674 TGGGAGCAGAAATTTGATAAATTTGACTTCGTCTTCAGGTGTAACTCTCCCAATCA 733
Db	598 TGTGGAGTGAATTCGATCGTTTGTGCTGTTTGGTGTAACTCTCCCAATCA 657
Qy	734 GGGAGCGCTAGTAAAGATGTTGTAAGCAAAACAAATCTTGTGACTGTCAATCCAGCAT 793
Db	658 GGAAGCATTTAGCAAGATGTCGCAATTAACAAACTTGTGACGCTTAATCCAAGTATC 717
Qy	794 ATAAACCTTGAAGTACCAGAATTTGAAGAGAAAGACACAGTTTTGGAGGACATCTCC 853
Db	718 ATAGCTCAGAAATACACAAACAAATGAAAAGAGACAGAAATTTGTGAGAACGTTGCA 777
Qy	854 ACCTATGGAGATGATCTCTCTCTGTCAGCATTTTCTATCGGCGCCCAACACAGGCATC 913
Db	778 GTCTATGGAGATGATCTCTCTCTTATGCGCGCATTTTCTTCCAGAAAGCAACACGCTACT 837
Qy	914 TCTTTTAAAGTCTACCAACACCAAGAGTCAAAAATGAGCAAAAAGGTTCTCTCTCTTC 973
Db	838 TCTTTTAAAGTCTATCACATTTSCAGGAGTTCAAGCAACCCAAAGGCAATATTTT 897
Qy	974 CATCCAGGTACTCTGAGACACCCCTCTTTTCTGGAGAACTAAAGGGTGTGACTGCATAC 1033
Db	898 CACCCACCTATCTGAAAATCTGCCCAGTTTGGAGAACTAAGGGTGTAAAGACCTAC 957
Qy	1034 CGTTTGTCCACAGGCTTGATGATGCAAGTGTGCTGTGGAACCTGTGTGAAAACCTGAAG 1093
Db	958 CGTTTATCATCTGTTTTATGACACTAGTCTGCTGTGTGAGTGTGTGAGAAATGTGAAG 1017
Qy	1094 CTCTACGATTTCTGCGCTTTCTTAAGACTATCGAAGACACCCCACTCAGTCACCACTAC 1153
Db	1018 CTTTACGGCTTCTGCGCTTTCTTAATCCACAGAGATGCCCCTAATCAGCCACCATAT 1077
Qy	1154 TATGATAACATGTTTACCTAAGCTGGTTTCCACAGATGCTCTTAAAGATATACAGCCAAATG 1213
Db	1078 TACGACAAACAGCTGCGCCCAACAGGTTTCCATGCAATGCCCCAGGAATACAAACAGATC 1137
Qy	1214 CTCAGCTCCATATGAGAGAACTCTCAACTGCNAATTCAGCAATGTGAAACGGTTAA 1273
Db	1138 CTTTCACTTTCATGTCGCAAGGCAATGAAAGTGTGCAATTTGGTAAATGTGAATCAGACTAA 1197
RESULT 15	
CR352448	520 bp mRNA linear VRT 11-MAR-2004
LOCUS	Gallus gallus finished cDNA, clone CHEST321n14.
DEFINITION	CR352448
ACCESSION	CR352448
VERSION	CR352448.1 GI:45422014
KEYWORDS	Gallus gallus (chicken)
SOURCE	Gallus gallus
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.
REFERENCE	1 (bases 1 to 1520)
AUTHORS	Boardman, P.E., Bonfield, W.K., Brown, W.R.A., Garder, C., Chalk, S.E., Croning, M.P.R., Davies, R.M., Francis, M.D., Grafham, D.V., Hubbard, S.J., Humphray, S.D., Hunt, P.J., Maddison, M., McLaren, S.R., Niblett, D., Overton, I.M., Rogers, J., Scott, C.E., Taylor, R.G., Tickle, C. and Wilson, S.A.
TITLE	Direct Submission
JOURNAL	Submitted (09-MAR-2004)
COMMENT	CB10 ISA, UK, E-mail enquiry: chibbs@bms.unist.ac.uk BBSRC/Dundee/Nottingham/anger/Sheffield/UMIST Gallus cdna sequencing project. This sequence is from the BBSRC/Dundee/Nottingham/anger/Sheffield/UMIST cdna collection, from a library constructed by Elizabeth Bosch. cdna was prepared from RNA extracted from spleen, normalised, and poly A-trimmed. ECoRI-NotI cut cdna was then ligated into the vector. Vector: pBluescript II KS(+); Site 1: EcoRI; Site 2: NotI Host: Escherichia coli DH10B.
FEATURES	Location/Qualifiers

source

1. .1520
/organism="Gallus gallus"
/mol_type="mRNA"
/strain="White Leghorn, Hisex"
/db_xref="taxon:9031"
/clone="ChEST32ln14"
/clone_lib="CSFQCHN24"
/dev_stage="stage 22"

ORIGIN

Query Match 14.4%; Score 456.8; DB 11; Length 1520;
Best Local Similarity 67.9%; Pred. No. 9.8e-138;
Matches 552; Conservative 0; Mismatches 307; Indels 1; Gaps 1;

Qy	314	AAATGCAAGATCTGCAATATAGCTTGAACTCTTTATCTAACAAACGAGACGGTACTCT	373
Db	229	AAATGTAAGGCAATTCAGGACCAACATTTTGTCTATCGTGGTTCAGGAAGAAAAGGTATCCA	288
Qy	374	GAGGATGACTACCTCCAGACCACCAACATACAGAGATGCCATGGAAACGGCAAGCA	433
Db	289	GAAGATTATTATCTTCACATAGTCAAAAACCTGCAGAACTGCACCTGGTAAAGGACCG	348
Qy	434	GAAGAAATGACAAATTTTAGACAAAACCTGGCTTCTCTGTTGGATGCCATTCAGACTTC	493
Db	349	GAAGAAATGACAAATTTTAG-TCAGAGTTAGCTTCTGCTGTAATGACAGTTTCAAAATTT	407
Qy	494	GTGGTTTCCCAAGAACCAACATCTCAGTGGGGACTAAACATGAGCTACGAGTGGAAAGCAAG	553
Db	408	ATTGCTTCTCAAAACCAACACCCCACTGGGAAGTAACATGAGTTATGAAGTGGACAGTAAA	467
Qy	554	AAACACATCCCACTTCGAGAGAACATTTTCCACATGTTTCCAGTGTGCGAGCCTTTTGTG	613
Db	468	AAGACCATCTCATTTACAGAGACATTTTGAATGCTGCCTGTGTCTCTCTCTTTCA	527
Qy	614	GACTATCCCTATAACAGTGTGAGTGTGTAATGGGGAATTCACAAAGTCTCTC	673
Db	528	GTTTATCCCATCAGAACTGTGAGTGTGTAATGGGGAATTCGAAATGGAGGATTTCTGAAAATTC	587
Qy	674	TGCGAGCAGAAATTTGATAATCTGACTTCGTCCTTCAGGTGTAACCTCCCGCCCAATACA	733
Db	588	TGTGAGCTGAAATCGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	647
Qy	734	GGGAGCGCTAGTAAAGATGTTGGAAGCAAAACAAATCTTGAGTGTCAATCCCGAGCAT	793
Db	648	GGAAGCATTTAGCAAGATGTCGCAATTAACAAACCTTTGACGGTTAATCCAAAGTATC	707
Qy	794	ATAACCTGAAGTACCAAGATTTGAAGGAGAGAAAGACAGTCTTTTGGAGGACATCTCC	853
Db	708	ATAGCTCAGAAATACAACAACTAAATGAAGAAGAGACAGAAATTTTGGAGAACGTTGCA	767
Qy	854	ACCTATGAGATGCAATCT	913
Db	768	GTCTATGAGATGCTTTCTTTTATTTGCGGCAATTTCTCTCAGAAGCAACACGCGCTACT	827
Qy	914	TCTTTTAAAGTCTACAAACACTCAAGAGTCAAAATATGAGGCAAAAGGTTCTCTCTCTC	973
Db	828	TCTTTTAAAGTCTATCACACTTTTGCAGGAGTTCAAAGCAACCCAAAGGGCAATATTTTT	887
Qy	974	CATCCAGGTACCTCAGACACCT	1033
Db	888	CACCCCACTATCTGAAATATCTTCCCAAGTTTGGAGAACTAAGGGTGTAAAGCCTAC	947
Qy	1034	CGCTTGTCCACAGGCTTGATGATGTCAGTGTGCTGTGGAACTGTGTGAAACGTTGAAG	1093
Db	948	CGGTTATCATCTGGTTTATGATCACTAGTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTG	1007
Qy	1094	CTCTACGATTTCTGCGCTTTCTTAAGACTATCAGAGACACCCCACTCAGTCCACCTAC	1153
Db	1008	CTTTACGGCTTCTGCGCTTTCTCAAAATCCACAGAGAGATGCCAATCAGCCACCACTAT	1067
Qy	1154	TATGATAACATGTTACGATGATGTTTCCACAGATGCTTAAAGATATACAGCCAAATG	1213
Db	1068	TACGACAAACAGCTGCCCCAACCGGTTTCCATGCAATGCCCAAGGAATACAAACAGATC	1127

Qy 1214 CTCAGCTCCATATGAGAGGAACTCAAACTGCAATTCAGCAATGTGAACGGCTTAA 1273
Db 1128 CTTAGCTTTCATGCGCAAGGCACTTTGAAGTTGCAAGTTGGTAAATGTGAATCAGACTAA 1187

Search completed: May 31, 2006, 15:48:56
Job time : 17619 secs

GenCore version 5.1.8
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 31, 2006, 10:49:29 ; Search time 14318 Seconds

(without alignments)
12364.898 Million cell updates/sec

Title: US-10-501-930-2

Perfect score: 3166

Sequence: 1 cggagcgcgagtcggtgcc.....gctattagcaaaaaaaaaa 3166

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 48236798 seqs, 2795965780 residues

Total number of hits satisfying chosen parameters: 96473596

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*

1: gb_est1:*

2: gb_est3:*

3: gb_est4:*

4: gb_est5:*

5: gb_est6:*

6: gb_hc:*

7: gb_est2:*

8: gb_est7:*

9: gb_est8:*

10: gb_est9:*

11: gb_ges1:*

12: gb_ges2:*

13: gb_ges3:*

14: gb_ges4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3068.6	96.9	3222	6 AK172400	Mus muscu
2	3026.2	95.6	3695	6 AK156557	Mus muscu
3	3010	95.1	3292	6 AK085105	Mus muscu
4	818	25.8	820	9 CX565877	UI-M-IBO-
5	760.4	24.0	812	2 BG919404	602818295
6	748.4	23.6	780	2 B1695458	603346462
7	718	22.7	753	2 B1653314	603301030
8	666.8	21.1	670	7 B4865599	BB486599
9	631	19.9	640	11 AZ315725	IM0032M24
10	624.8	19.7	628	7 B1112539	BB112539
11	604	19.1	625	7 B627781	BB627781
12	596.4	18.8	631	7 B627781	BB627781
13	575.4	18.2	663	4 B1752002	BY752002
14	573	18.1	852	2 B1653246	603300931
15	570	18.0	771	2 B1100002	602884742
16	482	15.2	517	7 B686184	uv73e09.x
17	468	14.8	936	4 CA976319	AGENCOURT
18	458	14.5	472	7 B678823	BB678823
19	456	14.4	463	7 B5553728	ur41e03.x

20	453.2	14.3	466	7 BB82306	BB832306
21	449.6	14.2	483	7 BB82306	BB829625
22	443.8	14.0	464	4 BY50796	BY50796
23	443.4	14.0	462	4 BY50796	BY50796
24	433.2	13.7	470	7 BB83042	BB83042
25	433.2	13.6	446	4 BY54759	BY54759
26	424.4	13.4	446	7 BB82186	BB82186
27	423.6	13.4	439	4 BY58373	BY58373
28	420.8	13.3	444	7 BB84183	BB84183
29	420	13.3	483	14 AY04993	AY04993
30	416	13.1	482	4 BY52845	BY52845
31	411.8	13.0	467	4 BY50610	BY50610
32	402.8	12.7	457	4 BY53730	BY53730
33	402.6	12.7	477	7 BB89600	BB89600
34	398	12.6	398	4 BY38484	BY38484
35	390.8	12.3	410	4 BY39395	BY39395
36	383.4	12.1	439	7 BB83290	BB83290
37	381	12.0	381	2 BF70564	mab73f04.
38	377.2	11.9	801	3 BU47532	BU47532
39	375.2	11.9	388	7 BB86534	BB86534
40	368.8	11.6	412	4 BY39548	BY39548
41	358.4	11.3	361	4 BY08633	BY08633
42	357	11.3	413	4 BY30069	BY30069
43	349.4	11.0	367	4 BY00379	BY00379
44	345.2	10.9	383	4 BY04148	BY04148
45	340.2	10.7	458	7 BB74521	BB74521

ALIGNMENTS

RESULT 1	AK172400	222 bp	mRNA	linear	HTC 21-SEP-2005
LOCUS	Mus musculus activated s	leen cDNA, RIKEN full-length enriched			
DEFINITION	8-sialyltransferase) F,	ull insert sequence.			
ACCESSION	AK172400				
VERSION	AK172400.1	GI:74186460			
KEYWORDS	HTC; CAP trapper.				
SOURCE	Mus musculus (house mouse)				
ORGANISM	Mus musculus				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muroidae; Muridae; Mus.				
AUTHORS	Carninci, P. and Hayashizaki, Y.				
TITLE	High-efficiency full-length cDNA cloning				
JOURNAL	Meth. Enzymol. 303, 19-4				
PUBMED	10349636				
REFERENCE	2				
AUTHORS	Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.				
TITLE	Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes				
JOURNAL	Genome Res. 10 (10), 1611-1630 (2000)				
PUBMED	11042159				
REFERENCE	3				
AUTHORS	Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwara, S., Inoue, K., Tanaka, Y., Izawa, M., Ohara, S., Watanabe, M., Yoneda, Y., Ishikawa, T., Sawa, K., Tanaka, T., Matsuura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.				
TITLE	RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multipillar sequencer				
JOURNAL	Genome Res. 10 (11), 1751-1771 (2000)				
PUBMED	11076861				
REFERENCE	4				
AUTHORS	Kawai, J., Shinagawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y., Arakawa, T., Hara, A., Fukunishi, Y., Konno, H., Adachi, J., Fukuda, S., Aizawa, K., Izawa, M., Nishikawa, K., Kiyosawa, H., Kondo, S., Yamanaka, I.,				

Saito, T., Okazaki, Y., Gojobori, T., Bono, H., Kasukawa, T., Saito, R., Kadota, K., Matsuda, H., Ashburner, M., Batalov, S., Casavant, T., Fleischmann, W., Gaasterland, T., Gissi, C., King, B., Kochiwa, H., Kuehl, P., Lewis, S., Matsumoto, Y., Nikaide, I., Pesole, G., Quackenbush, J., Schriml, L.M., Stabli, F., Suzuki, R., Tomita, M., Wagner, L., Washio, T., Sakai, K., Okido, T., Furuno, M., Aono, H., Baldarelli, R., Barsh, G., Blake, J., Boffelli, D., Bojunga, N., Carninci, P., de Bonaldo, M.F., Brownstein, M.J., Bult, C., Fletcher, C., Fujita, M., Gariboldi, M., Gustincich, S., Hill, D., Hofmann, L., Hume, D.A., Kamiya, M., Lee, N.H., Lyons, P., Marchionni, L., Mashima, J., Mazzarelli, J., Mombaerts, P., Nordone, P., Ring, B., Ringwald, M., Rodriguez, I., Sakamoto, N., Sasaki, H., Sato, K., Schonbach, C., Seys, T., Shibata, Y., Storch, K.F., Suzuki, H., Toyooka, K., Wang, K.H., Weitz, C., Whittaker, C., Wilming, L., Wynshaw-Boris, A., Yoshida, K., Hasegawa, Y., Kawaji, H., Kohtsuki, S. and Hayashizaki, Y.

Riken Genome Exploration Research Group Phase II Team and the PANTOM Consortium

Functional annotation of a full-length mouse cDNA collection
Nature 409 (6821), 685-690 (2001)
11217851

5

Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S., Nikaide, I., Osato, N., Saito, R., Suzuki, H., Yamanaka, I., Kiyosawa, H., Yagi, K., Tomaru, Y., Hasegawa, Y., Nogami, A., Schonbach, C., Gojobori, T., Baldarelli, R., Hill, D.P., Bult, C., Hume, D.A., Quackenbush, J., Schriml, L.M., Kanapin, A., Matsuda, H., Batalov, S., Beisel, K.W., Blake, J.A., Bradt, D., Bruscia, V., Chothia, C., Corbani, L.E., Cousins, S., Dalla, E., Dragani, T.A., Fletcher, C.F., Forrest, A., Frazer, K.S., Gaasterland, T., Gariboldi, M., Gissi, C., Godzik, A., Gough, J., Grimmond, S., Gustincich, S., Hirokawa, N., Jackson, I.J., Jarvis, E.D., Kanai, A., Kawaji, H., Kawasawa, Y., Kedzierski, R.M., King, B.L., Konagaya, A., Kurochkin, I.V., Lee, Y., Lenhard, B., Lyons, P.A., Maglott, D.R., Maltais, L., Marchionni, L., McKenzie, L., Miki, H., Nagashima, T., Numata, K., Okido, T., Pavan, W.J., Pertea, G., Pesole, G., Petrovsky, N., Pillai, R., Pontius, J.U., Qi, D., Ramachandran, S., Ravasi, T., Reed, J.C., Reid, D.J., Reid, J., Ring, B.Z., Ringwald, M., Sandelin, A., Schneider, C., Semple, C.A., Setou, M., Shimada, K., Sultana, R., Takenaka, Y., Taylor, M.S., Teasdale, R.D., Tomita, M., Verdaro, R., Wagner, L., Wahlestedt, C., Wang, Y., Watanabe, Y., Wells, C., Wilming, L.G., Wynshaw-Boris, A., Yanagisawa, M., Yang, I., Yang, L., Yuan, Z., Zavolan, M., Zhu, Y., Zimmer, A., Carninci, P., Hayatsu, N., Hirokawa, N., Kishikawa, T., Konno, H., Nakamura, M., Sakazume, N., Sato, K., Shiraki, T., Waki, K., Kawai, J., Aizawa, K., Arakawa, T., Fukuda, S., Hara, A., Hashizume, W., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Miyazaki, A., Sakai, K., Sasaki, D., Shibata, K., Shinagawa, A., Yasunishi, A., Yoshino, M., Waterston, R., Lander, E.S., Rogers, J., Birney, E. and Hayashizaki, Y.

PANTOM Consortium

Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
Nature 420 (6915), 563-573 (2002)
12466851

6

Carninci, P., Kasukawa, T., Katayama, S., Gough, J., Frith, M.C., Maeda, N., Oyama, R., Ravasi, T., Lenhard, B., Wells, C., Kodzius, R., Shimokawa, K., Bajic, V.B., Bremner, S.E., Batalov, S., Forrest, A.R., Zavolan, M., Davis, M.J., Wilming, L.G., Aidinis, V., Allen, J.E., Ambesi-Impombato, A., Apweiler, R., Aturaliya, R.N., Bailey, T.L., Banasik, M., Baxter, L., Belsel, K.W., Bersano, T., Bono, H., Chalk, A.M., Chiu, K.P., Choudhary, V., Christoffels, A., Clutterbuck, D.R., Crowe, M.L., Dalla, E., Dalrymple, B.P., de Bono, B., Della Gatta, G., di Bernardo, D., Down, T., Engstrom, P., Fagioli, M., Faulkner, G., Fletcher, C.F., Fukushima, T., Furuno, M., Fukui, S., Gariboldi, M., Georgii-Hemming, P., Gingeras, T., Gojobori, T., Green, R.E., Gustincich, S., Harbers, M., Hayashi, Y., Hensch, T.K., Hirokawa, N., Hill, D., Humineck, L., Iacono, M., Ikeo, K., Iwama, A., Ishikawa, T., Jakt, M., Kanapin, A., Katoh, M., Kawasawa, Y., Kelso, J., Kitamura, H., Kitano, H., Kollas, G., Krishnan, S.P., Kruger, A., Kummerfeld, S.K., Kurochkin, I.V., Lareau, L.F., Lazarevic, D., Lipovich, L., Liu, J., Liuni, S., McWilliam, S., Madan, Babu, M., Madera, M., Marchionni, L., Matsuda, H., Matsuzawa, S., Miki, H., Mignone, F., Miyake, S.,

Morris, K., Mottagui-Tabrizi, S., Mulder, N., Nakano, N., Nakauchi, H., Ng, P., Nilsson, R., Nishikuchi, S., Nishikawa, S., Nori, F., Ohara, O., Okazaki, Y., Orlando, V., Pang, C.C., Pavan, W.J., Pavese, J.F., Peolste, G., Petrovsky, N., Piazza, S., Reed, J.C., Reid, J.F., Ring, B.Z., Ringwald, M., Rost, B., Rubin, Y., Salzberg, S.L., Sandelin, A., Schneider, C., Schonbach, C., Sekiguchi, K., Semple, C.A., Seno, S., Sessa, L., Sheng, Y., Shibata, Y., Shimada, H., Shimada, K., Silva, D., Sinclair, B., Sperling, S., Stupka, E., Sugura, K., Sultana, R., Takenaka, Y., Taki, K., Tammo, J.A., Tan, S.L., Tang, S., Taylor, M.S., Tegner, J., Teichmann, S.A., Ueda, H.R., van Nimwegen, E., Verdaro, R., Wei, C.L., Yagi, K., Yamanashi, H., Zabarovsky, E., Zhu, S., Zimmer, A., Hume, D.A., Bult, C., Grimmond, S.M., Teasdale, R.D., Liu, E.T., Bruscia, V., Quackenbush, J., Wahlestedt, C., Mattick, J.S., Hume, D.A., Kal, C., Sasaki, D., Tomaru, Y., Fukuda, S., Kanamori-Katayama, M., Suzuki, M., Aoki, J., Arakawa, T., Iida, J., Imamura, K., Itoh, M., Kato, T., Kawaji, H., Kawashima, N., Kawashima, T., Kojima, M., Kondo, S., Konno, H., Nakai, B., Ninomiya, N., Nishio, T., Okada, M., Plessey, C., Shibata, K., Shiraki, T., Suzuki, S., Tagami, M., Waki, K., Watahiki, A., Okamura-Oho, Y., Suzuki, H., Kawai, J. and Hayashizaki, Y.

PANTOM Consortium

The transcriptome of the mammalian genome
Science 309 (5740), 1559-1563 (2005)
16141072

7

Katayama, S., Tomaru, Y., Kasukawa, T., Waki, K., Nakanishi, M., Nakamura, M., Nishida, H., Pap, C.C., Suzuki, M., Kawai, J., Suzuki, H., Carninci, P., Hayashizaki, K., Wells, C., Frith, M., Ravasi, T., Pang, K.C., Hallinan, J., Mattick, J., Hume, D.A., Lipovich, L., Batalov, S., Engstrom, P.G., Mizuno, Y., Faghini, M.A., Sandelin, A., Chalk, A.M., Mottagui-Tabrizi, S., Liang, Z., Lenhard, B. and Wahlestedt, C.

Riken Genome Exploration Research Group

Antisense transcription in the mammalian transcriptome
Science 309 (5740), 1564-1566 (2005)
16141073

8

(bases 1 to 3222)

Arakawa, T., Carninci, P., Fukuda, S., Hashizume, W., Hayashida, K., Hori, F., Iida, J., Imamura, K., Imotani, K., Itoh, M., Kanagawa, S., Kawai, J., Kojima, M., Konno, H., Murata, M., Nakamura, M., Ninomiya, N., Nishiyori, H., Nomura, K., Ohno, M., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shiraki, T., Tagami, M., Tagami, Y., Waki, K., Watahiki, A., Muramatsu, M. and Hayashizaki, Y.

Direct Submission

Submitted (14-APR-2004)

Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa, 230-0045, Japan (E-mail: genome-res@gs.riken.jp, URL: http://genome-gsc.riken.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)

cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.

Tissues were provided by Dr. John Todd (Dept. of Medical Genetics Wellcome Trust Centre for Molecular Mechanisms in Disease Wellcome Trust/MRC Building Addenbrookes Hospital Cambridge) whose assistance we gratefully acknowledge.

Please visit our web site for further details.

URL: http://genome-gsc.riken.jp/

URL: http://fantom.gsc.riken.jp/

Location/Qualifiers

1. 3222

/organism="Mus musculus"

/mol_type="mRNA"

/strain="NOD"

/db_xref="FANTOM DB:F830206C17"

/db_xref="taxon:10090"

/clone="F830206C17"

/tissue type="activated spleen"

/clone_lib="RIKEN full-length enriched mouse cDNA library"

142. 1359

CDS

/note="unnamed protein product; putative
sialyltransferase 8 (alpha-2, 8-sialyltransferase) F
(MG) [GI:2386797 8] [NM_145838, evidence: BLASTN, 98%,
match=3156] "
/protein_id="BAE42986.1"
/db_xref="GI:7418461"
/translation="MRSGTLPALIGSLMLLLMLCPADAPARSLMEGSRDIT
SGTSAALXLTSPPTPTPTPTSTYLDKQITQTEKCKDQVLSNLSNKTRYSEDD
YLOITNFORCPWNRQAEYDNFRKLASCDAIQDFVVSQNTPTNMSVEVESKK
HIPHRENFHMPFQSPVDYDYNQAVVNGGILNKLSCGABIDKSDVFRCLNLPPI
TSASDKVSTKLNLTNPSTIITLYQNLKKEKQFLIEDISTYDGFLLPAPFSTRAN
TGISFKVQTKLHKLKQVLFPHRYLHLALFWRTKGTAYRLSTGLMTASVAVEL
CENVKLYGFWPSPKSTIEDTPLSHYYDNMLPKHGFHPQKEYSQMLQLHMRGILKQF

Query Match		96.9%; Score 3068.6; DB 6; Length 3222;
Best Local Similarity		98.7%; Pred. No. 0;
Matches 3115; Conservative 0; Mismatches 39; Indels 3; Gaps 2;		
Qy	1	CGGAGCGGAGTCTGCGCCCGGCTGCGCTTCGCCCGGCGAGCTTTGGCGCGGAGG 60
Db	66	CGGAGCGGAGTCTGCGCCCGGCTGCGCTTCGCCCGGCGAGCTTTGGCGCGGAGG 125
Qy	61	ACGCCGCTGGCTCAGGATCAGATCGGGGGGACGCTGTTCGCCCTCATAGGCGAGCTGTAT 120
Db	126	ACGCCGCTGGCTCAGGATCAGATCGGGGGGACGCTGTTCGCCCTCATAGGCGAGCTGTAT 185
Qy	121	GCTGCTGCTCTCTGCGTATGCTCTGCTGCGCCGAGCGGCTGCGCGCTCCAGGCT 180
Db	186	GCTGCTGCTCTCTGCGTATGCTCTGCTGCGCCGAGCGGCTGCGCGCTCCAGGCT 245
Qy	181	GTTGATGAGGAGGAGCAGAGGACGACGCTGTGCTACCTCAGTGCACCTGAGACACTCTG 240
Db	246	GTTGATGAGGAGGAGCAGAGGACGACGCTGTGCTACCTCAGTGCACCTGAGACACTCTG 305
Qy	241	GAGCCCGACAAACCCCGGTACACGACGACGAGGACATATCTGGATGAGAAACACAC 300
Db	306	GAGCCCGACAAACCCCGGTACACGACGACGAGGACATATCTGGATGAGAAACACAC 365
Qy	301	CCAAATAACAGAGAAATGCAAGATCTGCAATATATAGCTTGAATCTTTTATCTAAACAA 360
Db	366	CCAAATAACAGAGAAATGCAAGATCTGCAATATATAGCTTGAATCTTTTATCTAAACAA 425
Qy	361	GAGCGGTACTCTGAGGATGACTACTCCAGACCATCAACATACAGAGATGCCCATG 420
Db	426	GAGCGGTACTCTGAGGATGACTACTCCAGACCATCAACATACAGAGATGCCCATG 485
Qy	421	GAAACCGGCAAGCAGAGAAATATGACAAATTTTACAGCAAAACCTGGCTTCTGTTGGATGC 480
Db	486	GAAACCGGCAAGCAGAGAAATATGACAAATTTTACAGCAAAACCTGGCTTCTGTTGGATGC 545
Qy	481	CATTCAAGACTCTGTTGTTTCCAGAACACATCCAGTGGGACATAACATGAGTACGA 540
Db	546	CATTCAAGACTCTGTTGTTTCCAGAACACATCCAGTGGGACATAACATGAGTACGA 605
Qy	541	GGTGAAGAGCAGAAACACATCCCATTCGAGAGAACATTTTCCACATGTTTCCAGTGC 600
Db	606	GGTGAAGAGCAGAAACACATCCCATTCGAGAGAACATTTTCCACATGTTTCCAGTGC 665
Qy	601	GCAGCTTTTGTGAGTACTCCCTATAACACAGTGTGCTGTTGTTGTAATGGGGAAATCT 660
Db	666	ACAGCTTTTGTGAGTACTCCCTATAACACAGTGTGCTGTTGTTGTAATGGGGAAATCT 725
Qy	661	CAACAAGTCTCTCTCGGAGCAGAAATTTGATAAATCTGACTTCGCTTTCAGGTGTAACCT 720
Db	726	CAACAAGTCTCTCTCGGAGCAGAAATTTGATAAATCTGACTTCGCTTTCAGGTGTAACCT 785
Qy	721	CCCCCAATCAAGGAGCGCTAGTAAAGATGTTGGAGCAAAACAAATCTTGTGACTGT 780
Db	786	CCCCCAATCAAGGAGCGCTAGTAAAGATGTTGGAGCAAAACAAATCTTGTGACTGT 845
Qy	781	CAATCCGAGCATTAACCTGAGTACGAGAAATTTGAGGAGAGAGACAGCTTTT 840

Db	846	CAATCCGAGCATTAATAACCCCTGAGTACCAGAAATTTGAAGGAGAGAAAGACACAGTTT 905
Qy	841	GGAGGACATCTCCACCTATGGAATGCAATCTCTCTCTCTGCGCAGCATTTTCTATCGGC 900
Db	906	GGAGGACATCTCCACCTATGGAATGCAATCTCTCTCTCTGCGCAGCATTTTCTATCGGC 965
Qy	901	CAACACAGCATCTCTTTTAAATCTACCAACACTCAAGAGAGTCAAAATGAGGCAAAA 960
Db	966	CAACACAGCATCTCTTTTAAATCTACCAACACTCAAGAGAGTCAAAATGAGGCAAAA 1025
Qy	961	GGTCTCTCTCTCCATCCAGGACCTGAGACACCTCGCTCTTTCTGAGAGAACTAAAG 1020
Db	1026	GGTCTCTCTCTCCATCCAGGACCTGAGACACCTCGCTCTTTCTGAGAGAACTAAAG 1085
Qy	1021	GGTGACTGCATACCGCTTGTCCAGGCTTGATGATGCAAGTGTGCTGTGGAAGTGTG 1080
Db	1086	GGTGACTGCATACCGCTTGTCCAGGCTTGATGATGCAAGTGTGCTGTGGAAGTGTG 1145
Qy	1081	TGAAAACCTGAAGCTCTACGGAATCTGCGCTTCTTAAGACTATCGAAGACACCCCACT 1140
Db	1146	TGAAAACCTGAAGCTCTACGGAATCTGCGCTTCTTAAGACTATCGAAGACACCCCACT 1205
Qy	1141	CAGTCAACACTATATGATTAACGTTTACCTAAGCATGGTTTCCACAGATGCTTAAAGA 1200
Db	1206	CAGTCAACACTATATGATTAACGTTTACCTAAGCATGGTTTCCACAGATGCTTAAAGA 1265
Qy	1201	ATACAGCCAAATGCTCCAGCTCATATGAGAGGATCTCAAACTGCAATTCAGCAAAATG 1260
Db	1266	ATACAGCCAAATGCTCCAGCTCATATGAGAGGATCTCAAACTGCAATTCAGCAAAATG 1325
Qy	1261	TGAAAACCTGAAGCTTAAAGCTTCTTAAGGAGATTAATTTTCAAGAGAGTGGAGTGTGTC 1320
Db	1326	TGAAAACCTGAAGCTTAAAGCTTCTTAAGGAGATTAATTTTCAAGAGAGTGGAGTGTGTC 1385
Qy	1321	CAGCATCTCAAAAAGCCAAATTAAGAGGACACAGAGAAAGCATGAATTAACAAAGCGCT 1380
Db	1386	CAGCATCTCAAAAAGCCAAATTAAGAGGACACAGAGAAAGCATGAATTAACAAAGCGCT 1445
Qy	1381	CTCCCACTGTCTAGACCAAGCACCCGCCCACTCACTTTGCGAGCTCCACAGTCA 1440
Db	1446	CTCCCACTGTCTAGACCAAGCACCCGCCCACTCACTTTGCGAGCTCCACAGTCA 1505
Qy	1441	CTCATCTCACTTCAACGTTCTTCTCTGAGATGAGACCAAAACATCAGACTTGGAT 1500
Db	1506	CTCATCTCACTTCAACGTTCTTCTCTGAGATGAGACCAAAACATCAGACTTGGAT 1565
Qy	1501	AAGTAAATGAGATTAATTTTCAATCATATGAAATTTGATTTGAGC--AGGGTCTCT 1558
Db	1566	AAGTAAATGAGATTAATTTTCAATCATATGAAATTTGATTTGAGC--AGGGTCTCT 1625
Qy	1559	CAGATGCTTCTTGTTCCTATCATGATAGCCATTCGCCCTTTTATCAGAGTGGTAAAG 1618
Db	1626	CAGATGCTTCTTGTTCCTATCATGATAGCCATTCGCCCTTTTATCAGAGTGGTAAAG 1685
Qy	1619	AACTGTCAATTTGCGCAAGCCCTTTCTGAGAGAAATGTCTGAATCATCGCGCGAGT 1678
Db	1686	AACTGTCAATTTGCGCAAGCCCTTTCTGAGAGAAATGTCTGAATCATCGCGCGAGT 1745
Qy	1679	TTTTCACACAGCTCTTCTCTTAAATAAATCTCTCCATCTCTCTCTAGTAGAGTA 1738
Db	1746	TTTTCACACAGCTCTTCTCTTAAATAAATCTCTCCATCTCTCTCTAGTAGAGTA 1805
Qy	1739	CAGAAAACAAATACCTTGTATGTTTTCAGGAGAAAGTCTTTTCTTAGCAATGTGCC 1798
Db	1806	CAGAAAACAAATACCTTGTATGTTTTCAGGAGAAAGTCTTTTCTTAGCAATGTGCC 1865
Qy	1799	TGCTTCTGATTCAGTTGCTTGGTGGGTTTGGGTTTGGTGGTGGTGGTGGTGGTGG 1858
Db	1866	TGCTTCTGATTCAGTTGCTTGGTGGGTTTGGGTTTGGTGGTGGTGGTGGTGGTGG 1924
Qy	1859	CGCTTCTTCACTTCTTCTTGTCTATATTTCTTACCTTTATCAGTTTGTATTCAGCTT 1918
Db	1925	CGCTTCTTCACTTCTTCTTGTCTATATTTCTTACCTTTTATCAGTTTGTATTCAGCTT 1984

TITLE JOURNAL PUBMED REFERENCE AUTHORS	Functional annotation of a full-length mouse cDNA collection		CONSRMT TITLE JOURNAL PUBMED REFERENCE AUTHORS	Watahiki, A., Okamura-Ono, Y., Suzuki, H., Kawai, J. and Hayashizaki, Y.			
	Nature 409 (6821), 685-690 (2001)	11217851		The transcriptional landscape of the mammalian genome	Science 309 (5740), 1559-1563 (2005)		
TITLE JOURNAL PUBMED REFERENCE AUTHORS	Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S., Nikaide, I., Oato, N., Saito, R., Suzuki, H., Yamanaka, I., Kiyosawa, H., Yagi, K., Tomaru, Y., Hasegawa, Y., Nogami, A., Schonbach, C., Gojobori, T., Baldarelli, R., Hill, D. P., Bult, C., Hume, D. A., Quackenbush, J., Schriml, L. M., Kanapin, A., Matsuda, H., Batalov, S., Beisel, K. W., Blake, J. A., Bradt, D., Brusic, V., Chochia, C., Corbani, L. E., Cousins, S., Dalla, E., Dragani, T. A., Fletcher, C. F., Forrest, A., Frazer, K. S., Gaasterland, T., Gariboldi, M., Gissi, C., Godzik, A., Gough, J., Grimmond, S., Gustincich, S., Hirokawa, N., Jackson, I. J., Jarvis, E. D., Kanai, A., Kawaji, H., Kawasawa, Y., Kedzierski, R. M., King, B. L., Konagaya, A., Kurochkin, I. V., Lee, Y., Lenhard, B., Lyons, P. A., Maglott, D. R., Maltica, L., Marchionni, L., McKenzie, L., Miki, H., Nagashima, T., Numata, K., Okido, T., Pavan, W. J., Perte, G., Pesole, G., Petrovsky, N., Pillai, R., Pontius, J. U., Qi, D., Ramachandran, S., Ravasi, T., Reed, J. C., Reed, D. J., Reid, J., Ring, B. Z., Ringwald, M., Sandelin, A., Schneider, C., Sempile, C. A., Setou, M., Shimada, K., Sultana, R., Takenaka, Y., Taylor, M. S., Teasdale, R. D., Tomita, M., Verardo, R., Wagner, L., Wahlestedt, C., Wang, Y., Watanabe, Y., Wells, C., Wilming, L. G., Wynshaw-Boris, A., Yanagisawa, M., Yang, I., Yang, L., Yuan, Z., Zavolan, M., Zhu, Y., Zimmer, A., Carninci, P., Hayatsu, N., Hirozane-Kishikawa, T., Konno, H., Nakamura, M., Sakazume, N., Sato, K., Shiraki, T., Waki, K., Kawai, J., Aizawa, K., Arakawa, T., Fukuda, S., Hara, A., Hashizume, W., Imotani, K., Iehii, Y., Itoh, H., Kagawa, I., Miyazaki, A., Sakai, K., Sasaki, D., Shibata, K., Shinagawa, A., Yasunishi, A., Yoshino, M., Waterston, R., Lander, E. S., Rogers, J., Birney, E. and Hayashizaki, Y.						
	PANTOM Consortium			Antisense transcription in the mammalian transcriptome	Science 309 (5740), 1566 (2005)		
TITLE JOURNAL PUBMED REFERENCE AUTHORS	Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs			Antisense transcription in the mammalian transcriptome	Science 309 (5740), 1566 (2005)		
	Nature 420 (6915), 563-573 (2002)	12466851		8 (bases 1 to 3695)			
TITLE JOURNAL PUBMED REFERENCE AUTHORS	Carninci, P., Kasukawa, T., Katayama, T., Gough, J., Frith, M. C., Maeda, N., Oyama, R., Ravasi, T., Lenhard, B., Wells, C., Kodzius, R., Shimokawa, K., Bajic, V. B., Brenner, S. E., Batalov, S., Forrest, A. R., Zavolan, M., Davis, M. J., Wilming, L. G., Aidinis, V., Allen, J. E., Ambesi-Impombato, A., Apweiler, R., Aturaliya, R. N., Bailey, T. L., Banal, M., Baxter, L., Beisel, K. W., Bersano, T., Bono, H., Chalk, A. M., Chiu, K. P., Choudhary, V., Christoffels, A., Clutterbuck, D. R., Crowe, M. L., Dalla, E., Dalrymple, B. F., de Bono, B., Della Gatta, G., Di Bernardo, D., Down, T., Engstrom, P., Fagioli, M., Faulkner, G., Fletcher, C. F., Fukushima, T., Furuno, M., Futaki, S., Gariboldi, M., Georgii-Hemming, P., Gingeras, T. R., Gojobori, T., Green, R. E., Gustincich, S., Harbers, M., Hayashi, Y., Hensch, T. K., Hirokawa, N., Hill, D., Humnicki, L., Iacono, M., Ikeo, K., Iwama, A., Ishikawa, T., Jakt, M., Kanapin, A., Katoh, M., Kawasawa, Y., Kelso, J., Kitamura, H., Kitano, H., Kollias, G., Krishnan, S. P., Kruger, A., Kummerfeld, S. K., Kurochkin, I. V., Lareau, L. F., Lazarevic, D., Lipovich, L., Liu, J., Liuni, S., McWilliam, S., Madan, Babu, M., Madera, M., Marchionni, L., Matsuda, H., Matsuzawa, S., Miki, H., Mignone, F., Miyake, S., Moris, K., Mottagui-Tabar, S., Mulder, N., Nakano, N., Nakauuchi, H., Ng, P., Nilsson, R., Nishiguchi, S., Nishikawa, S., Nori, P., Ohara, O., Okazaki, Y., Orlando, V., Pang, K. C., Pavan, W. J., Pavese, G., Pesole, G., Petrovsky, N., Piazza, S., Reed, J., Reid, J. F., Ring, B. Z., Ringwald, M., Rost, B., Ruan, Y., Salzberg, S. L., Sandelin, A., Schneider, C., Schonbach, C., Sekiguchi, K., Sempile, C. A., Seno, S., Sessa, Y., Sheng, Y., Shibata, Y., Shimada, H., Shimada, K., Silva, D., Sinclair, B., Sperling, S., Stupka, E., Sugtara, K., Sultana, R., Takenaka, Y., Taki, K., Tammoja, K., Tan, S. L., Tang, S., Taylor, M. S., Tegner, J., Teichmann, S. A., Ueda, H. R., van Nimwegen, E., Verardo, R., Wei, C. L., Yagi, K., Yamanishi, H., Zabarovsky, E., Zhu, S., Zimmer, A., Hide, W., Bult, C., Grimmond, S. M., Teasdale, R. D., Liu, E. T., Brusic, V., Quackenbush, J., Wahlestedt, C., Mattick, J. S., Hume, D. A., Kall, C., Sasaki, D., Tomaru, Y., Fukuda, S., Kanamori, Katayama, M., Suzuki, M., Aoki, J., Arakawa, T., Iida, J., Imamura, K., Itoh, M., Kato, T., Kawaji, H., Kawagashira, N., Kawashima, T., Kojima, M., Kondo, S., Konno, H., Nakano, K., Ninomiya, N., Nishio, T., Okada, M., Pleassy, C., Shibata, K., Shiraki, T., Suzuki, S., Tagami, M., Waki, K.,						
				Submitted (30-MAR-2004)			
TITLE JOURNAL PUBMED REFERENCE AUTHORS	Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa, 230-0045, Japan (E-mail: genome-res@gsr.riken.jp, URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)			Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa, 230-0045, Japan (E-mail: genome-res@gsr.riken.jp, URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)			
				Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa, 230-0045, Japan (E-mail: genome-res@gsr.riken.jp, URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)			
TITLE JOURNAL PUBMED REFERENCE AUTHORS	Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa, 230-0045, Japan (E-mail: genome-res@gsr.riken.jp, URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)			Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa, 230-0045, Japan (E-mail: genome-res@gsr.riken.jp, URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)			
				Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa, 230-0045, Japan (E-mail: genome-res@gsr.riken.jp, URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)			
TITLE JOURNAL PUBMED REFERENCE AUTHORS	Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa, 230-0045, Japan (E-mail: genome-res@gsr.riken.jp, URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)			Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa, 230-0045, Japan (E-mail: genome-res@gsr.riken.jp, URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)			
				Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa, 230-0045, Japan (E-mail: genome-res@gsr.riken.jp, URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)			
TITLE JOURNAL PUBMED REFERENCE AUTHORS	Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa, 230-0045, Japan (E-mail: genome-res@gsr.riken.jp, URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)			Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa, 230-0045, Japan (E-mail: genome-res@gsr.riken.jp, URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)			
				Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa, 230-0045, Japan (E-mail: genome-res@gsr.riken.jp, URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)			
TITLE JOURNAL PUBMED REFERENCE AUTHORS	Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa, 230-0045, Japan (E-mail: genome-res@gsr.riken.jp, URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)			Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa, 230-0045, Japan (E-mail: genome-res@gsr.riken.jp, URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)			
				Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa, 230-0045, Japan (E-mail: genome-res@gsr.riken.jp, URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)			
TITLE JOURNAL PUBMED REFERENCE AUTHORS	Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa, 230-0045, Japan (E-mail: genome-res@gsr.riken.jp, URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)			Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa, 230-0045, Japan (E-mail: genome-res@gsr.riken.jp, URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)			
				Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa, 230-0045, Japan (E-mail: genome-res@gsr.riken.jp, URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)			
TITLE JOURNAL PUBMED REFERENCE AUTHORS	Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa, 230-0045, Japan (E-mail: genome-res@gsr.riken.jp, URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)			Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa, 230-0045, Japan (E-mail: genome-res@gsr.riken.jp, URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)			
				Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa, 230-0045, Japan (E-mail: genome-res@gsr.riken.jp, URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)			
TITLE JOURNAL PUBMED REFERENCE AUTHORS	Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa, 230-0045, Japan (E-mail: genome-res@gsr.riken.jp, URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)			Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa, 230-0045, Japan (E-mail: genome-res@gsr.riken.jp, URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)			
				Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa, 230-0045, Japan (E-mail: genome-res@gsr.riken.jp, URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)			
TITLE JOURNAL PUBMED REFERENCE AUTHORS	Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa, 230-0045, Japan (E-mail: genome-res@gsr.riken.jp, URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)			Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa, 230-0045, Japan (E-mail: genome-res@gsr.riken.jp, URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)			
				Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa, 230-0045, Japan (E-mail: genome-res@gsr.riken.jp, URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)			
TITLE JOURNAL PUBMED REFERENCE AUTHORS	Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa, 230-0045, Japan (E-mail: genome-res@gsr.riken.jp, URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)			Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa, 230-0045, Japan (E-mail: genome-res@gsr.riken.jp, URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)			
				Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa, 230-0045, Japan (E-mail: genome-res@gsr.riken.jp, URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)			
TITLE JOURNAL PUBMED REFERENCE AUTHORS	Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa, 230-0045, Japan (E-mail: genome-res@gsr.riken.jp, URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)			Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa, 230-0045, Japan (E-mail: genome-res@gsr.riken.jp, URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)			
				Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa, 230-0045, Japan (E-mail: genome-res@gsr.riken.jp, URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)			
TITLE JOURNAL PUBMED REFERENCE AUTHORS	Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa, 230-0045, Japan (E-mail: genome-res@gsr.riken.jp, URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)			Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa, 230-0045, Japan (E-mail: genome-res@gsr.riken.jp, URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)			
				Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa, 230-0045, Japan (E-mail: genome-res@gsr.riken.jp, URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)			
TITLE JOURNAL PUBMED REFERENCE AUTHORS	Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa, 230-0045, Japan (E-mail: genome-res@gsr.riken.jp, URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)			Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa, 230-0045, Japan (E-mail: genome-res@gsr.riken.jp, URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)			
				Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa, 230-0045, Japan (E-mail: genome-res@gsr.riken.jp, URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)			
TITLE JOURNAL PUBMED REFERENCE AUTHORS	Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa, 230-0045, Japan (E-mail: genome-res@gsr.riken.jp, URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)			Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa, 230-0045, Japan (E-mail: genome-res@gsr.riken.jp, URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)			
				Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa, 230-0045, Japan (E-mail: genome-res@gsr.riken.jp, URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)			
TITLE JOURNAL PUBMED REFERENCE AUTHORS	Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa, 230-0045, Japan (E-mail: genome-res@gsr.riken.jp, URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)			Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa, 230-0045, Japan (E-mail: genome-res@gsr.riken.jp, URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)			
				Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa, 230-0045, Japan (E-mail: genome-res@gsr.riken.jp, URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)			
TITLE JOURNAL PUBMED REFERENCE AUTHORS	Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa, 230-0045, Japan (E-mail: genome-res@gsr.riken.jp, URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)			Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa, 230-0045, Japan (E-mail: genome-res@gsr.riken.jp, URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)			
				Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa, 230-0045, Japan (E-mail: genome-res@gsr.riken.jp, URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)			
TITLE JOURNAL PUBMED REFERENCE AUTHORS	Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa, 230-0045, Japan (E-mail: genome-res@gsr.riken.jp, URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)			Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa, 230-0045, Japan (E-mail: genome-res@gsr.riken.jp, URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)			
				Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa, 230-0045, Japan (E-mail: genome-res@gsr.riken.jp, URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)			
TITLE JOURNAL PUBMED REFERENCE AUTHORS	Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa, 230-0045, Japan (E-mail: genome-res@gsr.riken.jp, URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)			Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa, 230-0045, Japan (E-mail: genome-res@gsr.riken.jp, URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)			
				Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa, 230-0045, Japan (E-mail: genome-res@gsr.riken.jp, URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)			

Db	2815	AATTAAAACTAGTAATAAGAAACAAATCTTTAGGCCACATTAAGTGGGATCTTGAGTAAAAAGAA	2874
Qy	2335	AGGGAACAGCAGGAGAAAGCTGTTCCGCTTGGTTCTTGATTACCACAAATGAGCATCGCTGGAA	2394
Db	2875	AGGGAACAGCAGGAGAAAGCTGTTCCGCTTGGTTCTGATTACCCAAATGAGCATCGCTGGAA	2934
Qy	2395	GGAGGTTGTGAGGCTACCGCTAAACCTCTCGGTGAGGAGAGAGTACAGTGCATGAGTGTG	2454
Db	2935	GGAGGTTGTGAGGCTACCGCTAAACCTCTCTGTGTAGGAGAGAGTACAGTGCATGAGTGTG	2994
Qy	2455	GCGGCTTTTGTCCACACTCGCTGAAGGGTGAGTAATTCAGAGCCAAATCACATCACAAGGAT	2514
Db	2995	ACGGCTTTTGTCCACACTCGGTGAGGGTGAGTAATTCAGAGCCAAATCACATCACAAGGAT	3054
Qy	2515	GGACACACTTAACATCATCACTTCACAGGGGAGATGAATGCTTTTCATGAGAAATTAACACTCA	2574
Db	3055	GGACACACTTAACATCATCACTTCACAGGGGAGATGAAGCTTTTCATGAGAAATTAACACTCA	3114
Qy	2575	TAAGCTAAGCATCAGTTTGTAGTAAGAAATTTTCAGTAGAGTGTAAATATGAACATTTATATAC	2634
Db	3115	TAAGCTAAGCATCAGTTTGTAGTAAGAAATTTTCAGTAGAGTGTAAATATGAATCATTTATACAC	3174
Qy	2635	CTCTTACTAATGTGCCACCGACACCTTTTAATGTAAGCACATTTATTTATTAAGTGTACTT	2694
Db	3175	CTCTTACTAATGTCTCACTGACACCTTTTAATGTAAGCACATTTATTTGTGTTAGTATTTT	3234
Qy	2695	GACATTAATGCTTATGTCTGTATATCTGTGTTCAATCATCGATTTTCCCAAAAGTAAGA	2754
Db	3235	GATATTTACAAGCTTATGTCTGTATATCTGTGTTCAATCAAATTTTCCCAAAAGTAAGA	3294
Qy	2755	GCATAGGAGATGAGGCTTACATGACCAAGAAAACCTATAAATTTTACTCTTTAACTCTTACT	2814
Db	3295	GCATAGGAGATGAGGCCACATGCCAAGAAAACCTATAAATTTTACTGTTTAACTCTTACT	3354
Qy	2815	TGAGCCAGCTTGTGTTTATCAAGTGCTTTTTTGAAGAGACAGCACCCCTGTGAATTTCTTC	2874
Db	3355	TGAGCCAGCTTGTGTTTATCAAGTGCTTTTTTGAAGAGACAGCACCCCTGTAAATTTCTTC	3414
Qy	2875	ATTCCTGATACAGTGCACCTTGTAATTTAAACATTTGTATCTGTTTCAAGTTTACATCTC	2934
Db	3415	ATTCCTGATACAGTGCACCTTGTAATTTAAACATTTGTAACTTTGTTTCAAGTTTACATCTC	3474
Qy	2935	TTTCAATCTTTTATAGCAAAATCAACAGTATTTAGCTTCCAGAAAATTTATCAGAAGTTCATAT	2994
Db	3475	TTTCAATCTTTTATAGCAAAATCAACAGTATTTAGCTTCCAGAAAATTTATCAGAAGTTCATAT	3534
Qy	2995	ATAAATATTTTGCAAGGGTAAAAGGCTTTTTTGTGTTAAATATAAATAAAATTTTATTTT	3054
Db	3535	ATAAATATTTTGCAAGGGTAAAAGGCTTTTTTGTGTTAAATATAAATAAAATTTTATTTT	3594
Qy	3055	CTTCTGATGAATAGAGGCTCTTTTATGCTGCTGCTAATGAACCTTAATTAGCTTTTAAATTA	3114
Db	3595	CTTCTGATGAATAGAGGCTCTTTTATGCTGCTGCTAATGAACCTTAATTAGCTTTTAAATTA	3654
Qy	3115	TCTCCTTAGCAACATTTGGTGACGTTTCAATCATGCTATTATAGC	3155
Db	3655	TCTCCTTAGCAACATTTGGTGACGTTTCAATCATGCTATTATAGC	3695

Direct submission JOURNAL	Submitted (16-APR-2002) Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa, 230-0045, Japan URL:http://genome.gsc.riken.jp/ Fax:81-45-503-9216)	Ooshihide Hayashizaki, The Institute of Research (RIKEN), Laboratory for Genome Exploration Research Group (GSC), 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa, 230-0045, Japan (E-mail:genome-resgscgsc.riken.jp, en.jp/, Tel:81-45-503-9222, and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. Please visit our web site URL:http://genome.gsc.riken.jp/	COMMENT
------------------------------	---	--	---------

1931 GTGCTGCTTCTGATTCAGTTCGCTTGACATTAAGCTGGGTGGGTTGGTTGGAT 1990
1854 TTGGGGCGTTCTTCTCACTCTTTTGTCTATATTTTCTTAATCCTTTATCAGTTTGTATTCG 1913
1991 TTGGGGCGTTCTTCTCACTCTTTTGTCTATATTTTCTTAATCCTTTATCAGTTTGTATTCG 2050
1914 AGCTTCTGCTTTGGGATTCGCAATTCCTCTCCACACGACGAGGATCACTCAATGCA 1973
2051 AGCTTCTGCTTTGGGATTCGCAATTCCTCTCCACACGACGAGGATCACTCAATGCA 2110
1974 TAAAGTAGTTTCAAAACATCCATTCCTCTCCTCAATGTTTATCCATAAAGTTACTCATCTCA 2033
2111 TAAAGTAGTTTCAAAACATCCATTCCTCTCCTCAATGTTTATCCATAAAGTTACTCATCTCA 2170
2034 TTTTATTTAAATATAGTGAACATCTACTTGATATCAGACCCGAGGACCATCTCCATTTGA 2093
2171 TTTTATTTAAATATAGTGAACATCTACTTGATATCAGACCCGAGGACCATCTCCATTTGA 2230
2094 GAATATGAAGATATTGTCTACTGCGAGAAAGCAGGTGTGCGCATTTAAATTGATATGATAC 2153
2231 GAATATGAAGATATTGTCTACTGCGAGAAAGCAGGTGTGCGCATTTAAATTGATATGATAC 2290
2154 CACAAGCATCATCTGCCAGTTTATGAACACAGTGTCTGAAAGGATCATAGACAGGGGTGT 2213
2291 CACAAGCATCATCTGCCAGTTTATGAACACAGTGTCTGAAAGGATCATAGACAGGGGTGT 2350
2214 TAAATCTGATCCAGTAGAATAAATCTTCACTAGTGTACTATTTTCAGGGAAGAGTTAATTTCA 2273
2351 TAAATCTGATCCAGTAGAATAAATCTTCACTAGTGTACTATTTTCAGGGAAGAGTTAATTTCA 2410
2274 CAATTTAAACTAGTAAATGAACCAATCTTAGGCACATTAAGTGTATCTGAGTAAAGA 2333
2411 CAATTTAAACTAGTAAATGAACCAATCTTAGGCACATTAAGTGTATCTGAGTAAAGA 2470
2334 AAGGGAACAGCAGGAGAAAGCTGTTCGCTGTGTTCTGATTACCAAAATGAGCATCTGGA 2393
2471 AAGGGAACAGCAGGAGAAAGCTGTTCGCTGTGTTCTGATTACCAAAATGAGCATCTGGA 2530
2394 AGGAGTTGTGAGGCTACGCTAAAACTCTGCTGAGGAGAGATGATGATGATGATGATGAT 2453
2531 AGGAGTTGTGAGGCTACGCTAAAACTCTGCTGAGGAGAGATGATGATGATGATGATGAT 2590
2454 GCGGCTTTTGTCCACACTCTGAGAGGAGTGAATTCAGAGCCATCATCATCAAGA 2513
2591 GCGGCTTTTGTCCACACTCTGAGAGGAGTGAATTCAGAGCCATCATCATCAAGA 2650
2514 TGGACACACCTTAACATCATCTTCAAGGGGAGATGAATGCTTTTATGAGAAATTTACATCT 2573
2651 TGGACACACCTTAACATCATCTTCAAGGGGAGATGAATGCTTTTATGAGAAATTTACATCT 2710
2574 ATAAGCTAAGCATCAGTTTGTAGTAAAAATTTGAGTAGATGTTTAAATATGAACATTTTATA 2633
2711 ATAGCTAAGCATCAGTTTGTAGTAAAAATTTGAGTAGATGTTTAAATATGAACATTTTATA 2770
2634 CCTCTTACTAATGTCGCCACCGACCTTTTATGTAAGCACATTTATTTATTTAAGTTACT 2693
2771 CCTCTTACTAATGTCGCCACCGACCTTTTATGTAAGCACATTTATTTATTTAAGTTACT 2830
2694 TGACATTAATGCTTATGCTGTATATCTGTTTATCTCATCATCATCTTCCAAAAAGTAG 2753
2831 TGACATTAATGCTTATGCTGTATATCTGTTTATCTCATCATCATCTTCCAAAAAGTAG 2890
2754 AGCATAGGAGATGAGGCTTACATGCAAGAAATCTATAAATTTTACTCTTTAATTTCTTAC 2813
2891 AGCATAGGAGATGAGGCTTACATGCAAGAAATCTATAAATTTTACTCTTTAATTTCTTAC 2950
2814 TTGAGCCAGCTTTGTTTATGTAAGTGTCTTTTGAAGACAGCAGCCCTGTGAATTTCTT 2873
2951 TTGAGCCAGCTTTGTTTATGTAAGTGTCTTTTGAAGACAGCAGCCCTGTGAATTTCTT 3010
2874 CATTTCTGATACAGTCTCACTTGTATTTAAACATTTGTAATGTTGTTTCAAGTTTACATCT 2933

3011 CATTCTGTACAGTCTCACCTTATTTAACATTTGTAATGTTGTTTCAAGTTTACATCT 3070
2934 CTTTCATTTCTTTTATAGCAAAATAAACGATTTAGCTTCAGAAATTTATCAGAAATTCATA 2993
3071 CTTTCATTTCTTTTATAGCAAAATAAACGATTTAGCTTCAGAAATTTATCAGAAATTCATA 3130
2994 TATAAATATTTTGCAGAGGTTAAGGCTTTTGTAAATAAATAAAATTTATTTATTT 3053
3131 TATAAATATTTTGCAGAGGTTAAGGCTTTTGTAAATAAATAAAATTTATTTATTT 3190
3054 TCTTCTGATGAATAGAGGCTCTTTATGCTGCTGAATGAACCTTAATTTAGCTTTAAAT 3113
3191 TCTTCTGATGAATAGAGGCTCTTTATGCTGCTGAATGAACCTTAATTTAGCTTTAAAT 3250
3114 ATCTCTAGCAACATTTGTCACATTTCAATCATGCTATTAGC 3155
3251 ATCTCTAGCAACATTTGTCACATTTCAATCATGCTATTAGC 3292

RESULT 4
CX565877
LOCUS
DEFINITION
UI-M-180-cum-m-15-0-UI.1 NIH-BMAP_180 Mus musculus cDNA clone
IMAGE:30945902 5', mRNA sequence.
ACCESSION
CX565877
VERSION
CX565877.1 GI:57592906
KEYWORDS
EST.
SOURCE
Mus musculus (house mouse)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
1. (bases 1 to 820)
NIH-MGC <http://mgc.ncbi.nlm.nih.gov/>.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
CONTACT: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: Dr. James Lin University of Iowa
cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Distribution information can be found at
<http://genome.uiowa.edu/distribution/mousefl.html>
This clone was contributed by the Brain Molecular Anatomy Project
(BMAP)
Seq primer: pYX-5.
Location/Qualifiers
1. 820
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6"
/db_xref="taxon:10090"
/clone="IMAGE:30945902"
/tissue_type="whole eye"
/dev_stage="newborn (1,5,15 days)"
/lab_host="DH10 (T1 phage resistant)"
/clone_lib="NIH-BMAP_180"
/note="Organ: Eye; Vector: pYX-Asc; Site: 1: EcoR I;
Site 2: Not I; The library was constructed according
to the protocol of Benton and Davis, 1978, p. 9.791-806,
1996. Denatured RNA was size fractionated on a 1% agarose
gel. First strand cDNA synthesis was primed with oligo-dT
primer containing a Not I site. Double strand cDNA was
size selected according to mRNA size fraction, ligated
with EcoR I adapter, digested with Not I and then cloned
directionally into pYX-Rec vector. The library tag
sequence located between the Not I site and the polyA tail
is AATTAATACG. This library was created for the University
of Iowa Brain Anatomy Project (BMAP); 'Gene Discovery in the
Developing Mouse Nervous System', supported by National
Institute of Mental Health (NIMH)."

FEATURES
source

ORIGIN

Query Match	25.8%;	Score 818;	DB 9;	Length 820;	
Best Local Similarity	99.8%;	Pred. No. 1.8e-200;			
Matches 818;	Conservative 0;	Mismatches 2;	Indels 0;	Gaps 0;	
QY	448	TTTTAGAGCAAACTGGCTTCTCTGTTGGATGCCATTCAGAGCTTCGTTGTTCCAGAA	507		
DB	1	TTTTAGAGCAAACTGGCTTCTCTGTTGGATGCCATTCAGAGCTTCGTTGTTCCAGAA	60		
QY	508	CAACACTCCAGTGGGAGCTAACATGAGCTACGAGGTGAAAGCAAGAAACACATCCCAT	567		
DB	61	CAACACTCCAGTGGGAGCTAACATGAGCTACGAGGTGAAAGCAAGAAACACATCCCAT	120		
QY	568	TCGAGAGAACATTTCCACATGTTTCCAGTGTGCGAGCGCTTTGTGGACTATCCCTATAA	627		
DB	121	TCGAGAGAACATTTCCACATGTTTCCAGTGTGCGAGCGCTTTGTGGACTATCCCTATAA	180		
QY	628	CCAGTGTGAGTGGTGGTAAATGGGGAATTTCTCAACAAAGTCTCTCTCGGAGCAGAAAT	687		
DB	181	CCAGTGTGAGTGGTGGTAAATGGGGAATTTCTCAACAAAGTCTCTCTCGGAGCAGAAAT	240		
QY	688	TGATAAATCTGACTTCTGCTTTCAAGTGTAACTCTCCCTCCCAATCACAGGAGCGCTAGTAA	747		
DB	241	TGATAAATCTGACTTCTGCTTTCAAGTGTAACTCTCCCTCCCAATCACAGGAGCGCTAGTAA	300		
QY	748	AGATGTTGGAAGCAAAACAAATCTTGTGACTGTCAATCCAGCATATTAACCTGAAGTA	807		
DB	301	AGATGTTGGAAGCAAAACAAATCTTGTGACTGTCAATCCAGCATATTAACCTGAAGTA	360		
QY	808	CCAGAAATTTGAAGGAGAGAAAGCACAGTTTTTGGAGGACATCTCCACTATGGAGATGC	867		
DB	361	CCAGAAATTTGAAGGAGAGAAAGCACAGTTTTTGGAGGACATCTCCACTATGGAGATGC	420		
QY	868	ATTCTCTCTCTGCGAGCATTTTCTATCGGGCCAAACAGGACATCTCTTTAAAGTCTA	927		
DB	421	ATTCTCTCTCTGCGAGCATTTTCTATCGGGCCAAACAGGACATCTCTTTAAAGTCTA	480		
QY	928	CCAAACACTCAAGAGTCAAAATGAGGCAAAAGGTTCTCTTCTTCCATCCAGGTACT	987		
DB	481	CCAAACACTCAAGAGTCAAAATGAGGCAAAAGGTTCTCTTCTTCCATCCAGGTACT	540		
QY	988	GAGACACTCGCTCTTTTCTGGAGAACTAAAGGGGTGACTGCATACCGCTTGTCACAGG	1047		
DB	541	GAGACACTCGCTCTTTTCTGGAGAACTAAAGGGGTGACTGCATACCGCTTGTCACAGG	600		
QY	1048	CTTGATGATTCAGAGTGTGCTGTGGAACTGTGTGAAACCTGTAAGCTCTACGGATTCTG	1107		
DB	601	CTTGATGATTCAGAGTGTGCTGTGGAACTGTGTGAAACCTGTAAGCTCTACGGATTCTG	660		
QY	1108	GCCTTTCTCTAAGACTATCGAAGACACCCCACTCAGTCACCACTACTATGATTAACATGTT	1167		
DB	661	GCCTTTCTCTAAGACTATCGAAGACACCCCACTCAGTCACCACTACTATGATTAACATGTT	720		
QY	1168	ACCTAAGCATGTTTCCACCATGCTTAAAGAACTACAGCCAAATGCTCCAGCTCCATAT	1227		
DB	721	ACCTAAGCATGTTTCCACCATGCTTAAAGAACTACAGCCAAATGCTCCAGCTCCATAT	780		
QY	1228	GAGAGGAATCTCTCAAACTGCAATTCAGCAAAATGTGAAACG	1267		
DB	781	GAGAGGAATCTCTCAAACTGCAATTCAGCAAAATGTGAAACG	820		
RESULT 5	BG919404	812 bp	mRNA	linear	EST 05-JUN-2001
LOCUS	602818295F1	NCI_CGAP_Mam6	Mus musculus	cdna	IMAGE:4947286 5',
DEFINITION	mRNA sequence.				
ACCESSION	BG919404				
VERSION	BG919404.1	GI:14299880			
KEYWORDS	EST.				
SOURCE	Mus musculus	(house mouse)			
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				

Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muroidae; Muridae; Murinae; Mus.
 1 (Bases 1 to 812)
 NIH-MGC http://mgc.ncbi.nlm.nih.gov/.
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 Contact: Robert Straube, Ph.D.
 Email: csapbs-remail.nih.gov
 Tissue Procurement: Jeffrey Green M.D.
 cDNA Library Preparation: Life Technologies, Inc.
 The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: LLAM10896 row: n column: 23
 High quality sequence spot: 795.
 Location/Qualifiers
 1..812
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="FVB/N"
 /db_xref="taxon:10090"
 /clone="IMAGE:447286"
 /sex="female, virgin"
 /tissue_type="filtrating ductal carcinoma"
 /dev_stage="5 months"
 /lab_host="DH10"
 /clone_lib="NCI CGAP Mam6"
 /note="Organism: primary; Vector: pCMV-SPORT6; Site 1: SalI; Site 2: NotI; Cloned unidirectionally. Primer: Oligo dt. Library constructed by Life Technologies. Investigator providing samples: Jeffrey Green, M.D., NIH"

ORIGIN

Query Match	24.0%;	Score 760.4;	DB 2;	Length 812;	
Best Local Similarity	98.6%;	Pred. No. 1.6e-185;			
Matches 767;	Conservative 0;	Mismatches 11;	Indels 0;	Gaps 0;	
QY	164	CTGCGCGCTCCAGCGCTGTTGA	223		
DB	33	CCGGAAGAATCCGGCTGTTGA	92		
QY	224	GCATGAGACACTCTGAGCC	283		
DB	93	GCATGAGACACTCTGAGCC	152		
QY	284	CTGGATGAGAGACAAACCCAA	343		
DB	153	CTGGATGAGAGACAAACCCAA	212		
QY	344	TCTTTATCTAAACAAAGAGAC	403		
DB	213	TCTTTATCTAAACAAAGAGAC	272		
QY	404	ATACAGAGTCCCATGGAAC	463		
DB	273	ATACAGAGTCCCATGGAAC	332		
QY	464	GCTTCTCTGTTCCGATGCCATTC	523		
DB	333	GCTTCTCTGTTCCGATGCCATTC	392		
QY	524	ACTAATCATGACTACGAGTGG	583		
DB	393	ACTAATCATGACTACGAGTGG	452		
QY	584	CACATGTTCCAGTGTCCAGC	643		
DB	453	CACATGTTCCAGTGTCCAGC	512		
QY	644	GGTATGGGGAAATCTCAACA	703		

```

Db      513  GGTAAATGGGGAAATCTCAACAGATCTCTCTCGGAGCAGAAATTGATAAATCTGACTTC 572
QY      704  GTCTTCAGGTGTAACCTCCCTCCCAATCAGAGGAGCGCTAGTAAAGATGTTGGAACAAA 763
Db      573  GTCTTCAGGTGTAACCTCCCTCCCAATCAGAGGAGCGCTAGTAAAGATGTTGGAACAAA 632
QY      764  ACAATATCTTGTGACTGCTCAATCCAGCATTATACCTGAAGTACCAGAAATTTGAAGGAG 823
Db      633  ACAATATCTTGTGACTGCTCAATCCAGCATTATACCTGAAGTACCAGAAATTTGAAGGAG 692
QY      824  AAGAAAGCAGTTTTTGGAGGACATCTCCACCTATGGAGATGCAATTCCTCTCTGSCCA 883
Db      693  AAGAAAGCAGTTTTTGGAGGACATCTCCACCTATGGAGATGCAATTCCTCTCTGSCCA 752
QY      884  GCATTTTCTATCGGGCCACACAGGATCTCTTTTAAAGTGTACCAAAACATCAAG 941
Db      753  GCATTTTCTATCGGGCCACACAGGATCTCTTTTAAAGTGTACCAAAACATCAAG 810

RESULT 6
BI695458
LOCUS   603346462F1 NCI_CGAP_Mam2 Mus musculus cDNA clone IMAGE:5374215 5',
DEFINITION
ACCESSION BI695458
VERSION   1
KEYWORDS  mRNA sequence.
SOURCE    Mus musculus (house mouse)
ORGANISM  Mus musculus

REFERENCE
AUTHORS   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
TITLE     Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
JOURNAL   Sciurognathi; Muridae; Muridae; Murinae; Mus.
COMMENT   NIH-MGC http://mgi.nci.nih.gov/.
          National Institutes of Health, Mammalian Gene Collection (MGC)
          Unpublished (1999)
          Contact: Robert Strausberg, Ph.D.
          Email: cgabbs-remail.nih.gov
          Tissue Procurement: Gilbert Smith, Ph.D.
          cDNA Library Preparation: Life Technologies, Inc.
          cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
          DNA Sequencing by: Incyte Genomics, Inc.
          Clone distribution: MGC clone distribution information can be
          found through the I.M.A.G.E. Consortium/LLNL at:
          http://image.llnl.gov
          Plate: LLAM11952 row: k column: 16
          High quality sequence stop: 780.

FEATURES
source    Location/Qualifiers
           1..780
            /organism="Mus musculus"
            /mol_type="mRNA"
            /strain="FVB/N-3"
            /db_xref="taxon:10090"
            /clone="IMAGE:5374215"
            /tissue_type="tumor, biopsy sample"
            /dev_stage="5 months"
            /lab_host="DH10B"
            /clone_lib="NCI CGAP Mam2"
            /note="Organ: mammary; Vector: pCMV-SPORT6; Site 1: SalI;
            Site 2: NotI; Cloned unidirectionally. Primer: Oligo dT.
            Library constructed by Life Technologies. Investigator
            providing samples: Gilbert Smith, NIH"

ORIGIN
Query Match      23.6%; Score 748.4; DB 2; Length 780;
Best Local Similarity 99.9%; Pred.No. 2.1e-182;
Matches 749; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      845  GACATCTCCACCTATGGAGATGCAATTCCTCTGCGGAGCATTTTCTCTATCGGCCAAC 904
Db      1    GACATCTCCACCTATGGAGATGCAATTCCTCTGCGGAGCATTTTCTCTATCGGCCAAC 60
QY      905  ACAGGCACTCTTTTAAAGTGTACCAACACATCAAGAGTCAAAAATGAGGCCAAAGGTT 964

```

```

Db      61  ACAGGCACTCTTTTAAAGTGTACCAACACATCAAGAGTCAAAAATGAGGCCAAAGGTT 120
QY      965  CTCTTCTTCCATCCCGAGTACCGAGACACCTCGCTCTTTTCTGGAGAACTAAAGGGGTG 1024
Db      121  CTCTTCTTCCATCCCGAGTACCGAGACACCTCGCTCTTTTCTGGAGAACTAAAGGGGTG 180
QY      1025  ACTGCATACCGCTTGTCCACAGCTTGATGATTGCAAGTGTGCTCTGGAACTGTGTGAA 1084
Db      181  ACTGCATACCGCTTGTCCACAGCTTGATGATTGCAAGTGTGCTCTGGAACTGTGTGAA 240
QY      1085  AACGTGAAGCTCTACGGATTCTGGCTTCTCTTAAGACTATCGAAGACACCCCACTCAGT 1144
Db      241  AACGTGAAGCTCTACGGATTCTGGCTTCTCTTAAGACTATCGAAGACACCCCACTCAGT 300
QY      1145  CACCACTACTATGATAACATGTACCTAAGCATGTTTCCACAGATGCTTAAAGAATAC 1204
Db      301  CACCACTACTATGATAACATGTACCTAAGCATGTTTCCACAGATGCTTAAAGAATAC 360
QY      1205  AGCCAAATGCTCCAGTCCATAGAGGAATCTCTCAAACTGCNAATTCAGCAATCTGAA 1264
Db      361  AGCCAAATGCTCCAGTCCATAGAGGAATCTCTCAAACTGCNAATTCAGCAATCTGAA 420
QY      1265  ACGGCTTAAAGTCTTCTTAGAAGCAGAAATAATTTCAAGAGGTGGAGTGGATGTGTACAGC 1324
Db      421  ACGGCTTAAAGTCTTCTTAGAAGCAGAAATAATTTCAAGAGGTGGAGTGGATGTGTACAGC 480
QY      1325  ATCTCCAAAAGCCCAATAGAAGAGGCACAGAGAAGCATGAATTAACAAGGCGCTCTCC 1384
Db      481  ATCTCCAAAAGCCCAATAGAAGAGGCACAGAGAAGCATGAATTAACAAGGCGCTCTCC 540
QY      1385  GACTTGTCTAGACCAAAAGCCACCGCCCCACTCTCTTTGCGAGCTCCAGGAGTCACTCA 1444
Db      541  CACTTGTCTAGACCAAAAGCCACCGCCCCACTCTCTTTGCGAGCTCCAGGAGTCACTCA 600
QY      1445  TTCTCACCTCTCAACGTTCTTTCTCTGAGAAATAGAGACCAAAACATCAGACTTGGATAAGT 1504
Db      601  TTCTCACCTCTCAACGTTCTTTCTCTGAGAAATAGAGACCAAAACATCAGACTTGGATAAGT 660
QY      1505  AAAATGAGATAAATTTTCAAAATATCATAGAAATTTGATTGAGCCAGGCTCTCTCAGAAT 1564
Db      661  AAAATGAGATAAATTTTCAAAATATCATAGAACTTTGATTGAGCCAGGCTCTCTCAGAAT 720
QY      1565  GCTTCTCTGTTCTATCCATGAGCCATT 1594
Db      721  GCTTCTCTGTTCTATCCATGAGCCATT 750

RESULT 7
BI653314
LOCUS   603301030F1 NCI_CGAP_Mam2 Mus musculus cDNA clone IMAGE:5341591 5',
DEFINITION
ACCESSION BI653314
VERSION   1
KEYWORDS  mRNA sequence.
SOURCE    Mus musculus (house mouse)
ORGANISM  Mus musculus

REFERENCE
AUTHORS   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
TITLE     Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
JOURNAL   Sciurognathi; Muridae; Muridae; Murinae; Mus.
COMMENT   NIH-MGC http://mgi.nci.nih.gov/.
          National Institutes of Health, Mammalian Gene Collection (MGC)
          Unpublished (1999)
          Contact: Robert Strausberg, Ph.D.
          Email: cgabbs-remail.nih.gov
          Tissue Procurement: Gilbert Smith, Ph.D.
          cDNA Library Preparation: Life Technologies, Inc.
          cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
          DNA Sequencing by: Incyte Genomics, Inc.
          Clone distribution: MGC clone distribution information can be
          found through the I.M.A.G.E. Consortium/LLNL at:
          http://image.llnl.gov
          Plate: LLAM11952 row: k column: 16
          High quality sequence stop: 780.

```

659 CACAACATCATCCAGTTTGAAACACAGTCGTGAAGGATCATAGACACAGGGTGGT 719

2214 TAAATCTGATCCAGTAGAATA 2236

719 TAAATCTGATCCAGTAGAATA 741

RESULT 8

BB486599

LOCUS

DEFINITION

BB486599 RIKEN full-length cDNA clone D430137E10 3', mRNA sequence. EST 25-OCT-2001

ACCESSION

BB486599

VERSION

BB486599.2

KEYWORDS

GI:16441407

SOURCE

Mus musculus (house mouse)

ORGANISM

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE

1 (bases 1 to 670)

Arakawa,T., Carninci,P., Hara,A., Hiramoto,K., Horiuchi,P., Ishii,Y., Ito,M., Kawai,J., Konno,H., Kouda,M., Koyama,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M., Okazaki,Y., Okimoto,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D., Shibata,K., Shingawa,A., Shiraki,T., Sugabe,Y., Suzuki,H., Takami,M., Tagawa,A., Takahashi,F., Takenaka,Y., Tanaka,T., Toyota,T., Muramatsu,M. and Hayashizaki,Y. RIKEN Mouse ESTs (Arakawa,T., et al. 2001)

AUTHORS

Unpublished (2001)

TITLE

On Jul 23, 2000 this sequence version replaced gi:9404208.

JOURNAL

Contact: Yoshihide Hayashizaki

COMMENT

Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute

The Institute of Physical and Chemical Research (RIKEN)

1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan

Tel: 81-45-503-9222

Fax: 81-45-503-9216

Email: genome-res@gsc.riken.jp, URL: http://genome.gsc.riken.jp/

Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y. Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)

wagi,K., Fujiwara,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M., Yoneda,Y., Shikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y. RIKEN integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)

Konno,H., Fukunishi,Y., Sugahara,Y. and Hayashizaki,Y. Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)

Kondo,S., Shingawa,A., Saito,T., Kiyosawa,H., Yamana,K., Aizawa,K., Fukuda,S., Hayashizaki,Y. and Hayashizaki,Y. Full-Length Mouse cDNAs Compared with Human Genome Sequences. Genome. 12, 673-677 (2001)

Please visit our web site (http://genome.gsc.riken.go.jp/) for further details.

cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.

Location/Qualifiers

1. .670

/organism="Mus musculus"

/mol_type="mRNA"

/db_xref="taxon:10090"

|||||
640 GAAGAGGATATTTTCAGAGAGTGAGTGATGTGTACAGCATCTCCAAAGCAATA 581
QY 1343 GAAGAAGCAGACAGAAAGCATGAATTAACAAGGCGCTCTCCCACTGTGTAGACCAAG 1402
Db 580 GAAGAAGCAGACAGAAAGCATGAATTAACAAGGCGCTCTCCCACTGTGTAGACCAAG 521
QY 1403 CCACCGGCGCCCACTCACTTTGACGCTCCAGAGTCACTCATTTTCACCTTCAAGTTC 1462
Db 520 CCACCGGCGCCCACTCACTTTGACGCTCCAGAGTCACTCATTTTCACCTTCAAGTTC 461
QY 1463 TTTCTCTGAGATAGAGACCAAAACATCAGACTTGGTAAGTAAATGAGATAATTTTTC 1522
Db 460 TTTCTCTGAGATAGAGACCAAAACATCAGACTTGGTAAGTAAATGAGATAATTTTTC 401
QY 1523 AAATCATCATGAATTTGATTTGAGCCAGGGTCTCTCAGAAATGCTTCTTCTATCC 1582
Db 400 AAATCATCATGAATTTGATTTGAGCCAGGGTCTCTCAGAAATGCTTCTTCTATCC 341
QY 1583 ATGATAGCCATTCACCTTTATCAGAGTGGTAAGTAAATGAGATAATTTGCGCAAGAC 1642
Db 340 ATGATAGCCATTCACCTTTATCAGAGTGGTAAGTAAATGAGATAATTTGCGCAAGAC 281
QY 1643 CTTTCTGAAGAGATGCTGAATCATGCGCGAGTTTTTACACAGACTTCTTCTTTAT 1702
Db 280 CTTTCTGAAGAGATGCTGAATCATGCGCGAGTTTTTACACAGACTTCTTCTTTAT 221
QY 1703 AAATAAATCTTCCCATTTCTCCTCTAGTAGAGTACAGAAACAAATACCTTGATGAT 1762
Db 220 AAATAAATCTTCCCATTTCTCCTCTAGTAGAGTACAGAAACAAATACCTTGATGAT 161
QY 1763 TCAGGAAGAAAGCTTTTCTTACTTAGCAATGCTCTGCTTCTGATTCAGTTCGTTTG 1822
Db 160 TCAGGAAGAAAGCTTTTCTTACTTAGCAATGCTCTGCTTCTGATTCAGTTCGTTTG 101
QY 1823 ACATTAAGCTGGTGGGGTTTGGTGGATTTGGGGGTTTCTTCACTTCTTTTGCTA 1882
Db 100 ACATTAAGCTGGTGGGGTTTGGTGGATTTGGGGGTTTCTTCACTTCTTTTGCTA 41
QY 1883 TATTTTCTTACCTTATCAGTTTATTCG 1913
Db 40 TATTTTCTTACCTTATCAGTTTATTCG 10

RESULT 10

BB112539
LOCUS
DEFINITION BB112539 RIKEN full-length enriched mouse cDNA library, C57BL/6J urinary bladder male adult Mus musculus cDNA clone 9530036E06 3', mRNA sequence.

ACCESSION

BB112539

VERSION

BB112539.2 GI:15408053

KEYWORDS

EST.

SOURCE

Mus musculus (house mouse)

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muridea; Muridae; Murinae; Mus.
1 (bases 1 to 628)
The FANTOM Consortium and RIKEN Genome Exploration Research Group and Genome Science Group (Genome Network Project Core Group).
The transcriptional landscape of the mammalian genome
Science 309 (5740), 1559-1563 (2005)
16141072

COMMENT

On Jun 27, 2000 this sequence version replaced gi:8765107.
Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gs.riken.jp, URL: http://genome.gsc.riken.jp/

cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.
Please visit our web site for further details.
URL: http://genome.gsc.riken.jp/.

FEATURES

source

1..628
/location/Qualifiers
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="9530036E06"
/sex="male"
/tissue_type="urinary bladder"
/dev_stage="adult"
/cDNA_lib="RIKEN full-length enriched mouse cDNA library, C57BL/6J urinary bladder male adult"

ORIGIN

Query Match 19.7%; Score 624.8; DB 7; Length 628;
Best Local Similarity 99.7%; Prd. No. 2.4e-150;
Matches 626; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1361 GCATGAATTACAAAGGCGCTCTCCACTTGTCTAGACCAAGGCCACCCGCCCACTCAC 1420
Db 1 GCATGAATTACAAAGGCGCTCTCCACTTGTCTAGACCAAGGCCACCCGCCCACTCAC 60
QY 1421 TTTGCAGCTCCACAGTCACTTATTTCTACCTTCAACGTTCTTTCTCTGAGAAATAGAGA 1480
Db 61 TTTGCAGCTCCACAGTCACTTATTTCTACCTTCAACGTTCTTTCTCTGAGAAATAGAGA 120
QY 1481 CCAAAACATCAGACTTGGATAAATAATGAGATAATTTTCAAATCATCATAGAAATTTG 1540
Db 121 CCAAAACATCAGACTTGGATAAATAATGAGATAATTTTCAAATCATCATAGAAATTTG 180
QY 1541 ATTTGAGCCAGGGTCTCTCAGATGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1600
Db 181 ATTTGAGCCAGGGTCTCTCAGATGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 240
QY 1601 TTTATCAGAGTGGTAATCAAAACGTGCAATTTGTGCCAAAGACCCCTTTCTGAAGAGAAATGT 1660
Db 241 TTTATCAGAGTGGTAATCAAAACGTGCAATTTGTGCCAAAGACCCCTTTCTGAAGAGAAATGT 300
QY 1661 CTGAATCATGCGCGAGTCTTTTACACAGCTCTTCTTATAAATAAATCTTCTTCTTCTTCT 1720
Db 301 CTGAATCATGCGCGAGTCTTTTACACAGCTCTTCTTATAAATAAATCTTCTTCTTCTTCT 360
QY 1721 CTCCTCTCTAGTAGAGTACAGAACAAATACCTTGTGATTCAGGAAGAAAGTCTTTT 1780
Db 361 CTCCTCTCTAGTAGAGTACAGAACAAATACCTTGTGATTCAGGAAGAAAGTCTTTT 420
QY 1781 TTTACTTAGCAATGTGCTGCTCTGATTTAGTTCAGTTTCGTTGTGACATTAAGCTGGGTGGG 1840
Db 421 TTTACTTAGCAATGTGCTGCTCTGATTTAGTTCGTTGTGACATTAAGCTGGGTGGG 480
QY 1841 GTTTTGGTGGATTTGGGGCGTCTTCTCATCTTTTGTCTATATTTTCTTCTTCTTCTTCT 1900
Db 481 GTTTTGGTGGATTTGGGGCGTCTTCTCATCTTTTGTCTATATTTTCTTCTTCTTCTTCT 540
QY 1901 CAGTTTCTGATTCGAGCTTCTCTGTTGGGATTCGCAATTCCTCTCTCCCACTGACAGAT 1960
Db 541 CAGTTTCTGATTCGAGCTTCTCTGTTGGGATTCGCAATTCCTCTCTCCCACTGACAGAT 600
QY 1961 CAACCTCAATGACATAAAGTAGTCAAAAC 1988
Db 601 CAACCTCAATGACATAAAGTAGTCAAAAC 628

RESULT 11

BB627781

LOCUS

BB627781

525 bp mRNA

linear

EST 08-NOV-2005

DEFINITION BB627781 RIKEN full-length enriched mouse cDNA library, C57BL/6J urinary bladder male adult Mus musculus cDNA clone 953003E06 5', mRNA sequence.

ACCESSION BB627781

VERSION BB627781.1 GI:15399224

KEYWORDS EST.

SOURCE Mus musculus (house mouse)

ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 625)
The PANTOM Consortium and RIKEN Genome Exploration Research Group and Genome Science Group (Genome Network Project Core Group). The transcriptional landscape of the mammalian genome Science 309 (5740), 1559-1563 (2005)

16141072
Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-Ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gs.c.riken.jp, URL: http://genome.gsc.riken.jp/
cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.
Please visit our web site for further details.
URL: http://genome.gsc.riken.jp/.

FEATURES
source
Location/Qualifiers
1..625
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="953003E06"
/sex="male"
/tissue_type="urinary bladder"
/dev_stage="adult"
/clone_lib="RIKEN full-length enriched mouse cDNA library, C57BL/6J urinary bladder male adult"

ORIGIN
Query Match 19.1%; Score 604; DB 7; Length 625;
Best Local Similarity 100.0%; Pred. No. 6e-145;
Matches 604; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CGGAGCGGAGTCGGTGGCGCGCGCGGCTGCGCTTCGCCCGCGAGCTTTGGCGGCGAGG 60
DB 21 CGGAGCGGAGTCGGTGGCGCGCGCGGCTGCGCTTCGCCCGCGAGCTTTGGCGGCGAGG 80
QY 61 AGCCCGCTGGCTCAGGATGAGTCCGGGGGACGCTGTTCGCCCTCATAGGCGAGCTGAT 120
DB 81 AGCCCGCTGGCTCAGGATGAGTCCGGGGGACGCTGTTCGCCCTCATAGGCGAGCTGAT 140
QY 121 GCTGCTGCTCTCTCGCTATGCTCTCTGTGTCGCCGAGCGGCTGCGCGCTCCAGGCT 180
DB 141 GCTGCTGCTCTCTCGCTATGCTCTGTGTCGCCGAGCGGCTGCGCGCTCCAGGCT 200
QY 181 GTTGATGAGGGAAGCAGAGGAGCACCACTGGTACCTCAGCTGCTGACCTGAAGACACTCTG 240
DB 201 GTTGATGAGGGAAGCAGAGGAGCACCACTGGTACCTCAGCTGCTGACCTGAAGACACTCTG 260
QY 241 GAGCCCGACACCCGGTACCGACGACGAGGACGACATATCTGGATGAGGAGACAC 300
DB 261 GAGCCCGACACCCGGTACCGACGACGAGGACGACATATCTGGATGAGGAGACAC 320
QY 301 CCAATAACAGAGAAATGCAAGATCTGCAATATAGCTTGAATCTTTATCTTAAACAAAC 360
DB 321 CCAATAACAGAGAAATGCAAGATCTGCAATATAGCTTGAATCTTTATCTTAAACAAAC 380

QY 361 GAGACGGTACTCTGAGGATGACCTCCAGACCATCACAACATACAGAGATGCCCATG 420
DB 381 GAGACGGTACTCTGAGGATGACCTCCAGACCATCACAACATACAGAGATGCCCATG 440
QY 421 GAACCGGCAAGCAGAGAGATATACAAATTTAGAGCAAACTGGCTTCCTGTTCGGATGC 480
DB 441 GAACCGGCAAGCAGAGAGATATACAAATTTAGAGCAAACTGGCTTCCTGTTCGGATGC 500
QY 481 CATTCGAAGCTTCGTGGTTCCTAGAAACAACTCCAGTGGGACCTAAACATGAGCTACGA 540
DB 501 CATTCGAAGCTTCGTGGTTCCTAGAAACAACTCCAGTGGGACCTAAACATGAGCTACGA 560
QY 541 GGTGGAAACCAAGAACACATCCCAATTCGAGAGACATTTCCACATGTTTCCAGTGTC 600
DB 561 GGTGGAAACCAAGAACACATCCCAATTCGAGAGACATTTCCACATGTTTCCAGTGTC 620
QY 601 GCAG 604
DB 621 GCAG 624

RESULT 12
BE633149/c
LOCUS BE633149
DEFINITION uv73e09.yi Soares mouse Ndms Mus musculus cDNA clone IMAGE:3412840 5' similar to TR:P97713 97713 GD3 SYNTHASE ;, mRNA sequence.

ACCESSION BE633149

VERSION BE633149.1

KEYWORDS EST.

SOURCE Mus musculus (house mouse)

ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 631)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs@mail.nih.gov
This clone is available royalty-free through LLNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:1088652

COMMENT Possible reversed clone: similarity on wrong strand
Seq primer: -40RP from G-bco
High quality sequence step: 467.

FEATURES
source
Location/Qualifiers
1..631
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:3412840"
/sex="male"
/tissue_type="Spleen"
/dev_stage="4 weeks"
/lab_host="DH10"
/clone_lib="Soares mouse 3Ndms"
/note="Vector: 7773D-PacI, Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5', TGTTACCAATCTGAATGGGAGCGCGCGCTGTTTTTTTTTTTTTTTTTTTTTTT 3']; double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI site of the modified p7773 vector. RNA provided by Dr. Bertrand Jordan. Library went through three rounds of normalization, and was constructed by Bento Soares and M. Fatima Bonaldo."

ORIGIN
Query Match 18.8%; Score 596.4; DB 7; Length 631;

Best Local Similarity 96.5%; Pred. No. 5.7e-143; Matches 609; Conservative 0; Mismatches 22; Indels 0; Gaps 0;	
QY 755 GGAAGCAAAACAAATCTTGTGACTGTCAATCCAGCATTAATACCTCGAATACAGAAAT 814	
Db 631 GGAAGCAAAACAAATCTTGTGACTGTCAATCCAGCAATTAATACCTCGAATACAGAAAT 572	
QY 815 TTGAAGGAGGAAGAACACAGTTTGTGGAGGACATCTCCACCTATGGAGATCATCTCTTC 874	
Db 571 TTGAAGGAGGAAGAACACAGTTTGTGGAGGACATCTCCACCTATGGAGATCATCTCTTC 512	
QY 875 CTCCTGCCAGCATTTTCTTATCGGGCAACACAGGATCTCTTTTAAAGTCTACCAACA 934	
Db 511 CTCCTGCCAGCATTTTCTTATCGGGCAACACAGGATCTCTTTTAAAGTCTACCAACA 452	
QY 935 CTCAAAGAGTCAAAATGAGCAAAAGGTTCTTTCTCCATCCAGGTACCTGAGACAC 994	
Db 451 CTCAAAGAGTCAAAATGAGCAAAAGGTTCTTTCTCCATCCAGGTACCTGAGACAC 392	
QY 995 CTCGCTCTTTCTGGAGAACTAAAGGGTGACTGCATACCGCTTGTCCACAGGCTTGATG 1054	
Db 391 CTCGCTCTTTCTGGAGAACTAAAGGGTGACTGCATACCGCTTGTCCACAGGCTTGATG 332	
QY 1055 ATTGCAAGTGTGCTGTGGAACGTGTGTAACAGTGAAGCTCTACGGATCTTGGGCTTTC 1114	
Db 331 ATTGCAAGTGTGCTGTGGAACGTGTGTAACAGTGAAGCTCTACGGATCTTGGGCTTTC 272	
QY 1115 TCTAAGACTATCGAAGACACCCCTCAGTCACTACTATGATTAACATGTTACTTAAG 1174	
Db 271 TCTAAGACTATCGAAGACACCCCTCAGTCACTACTATGATTAACATGTTACTTAAG 212	
QY 1175 CATGCTTTCCACAGATGCTTAAGAAATACAGCAAAAGTCTCCAGCTCCATATGAGGA 1234	
Db 211 CATGCTTTCCACAGATGCTTAAGAAATACAGCAAAAGTCTCCAGCTCCATATGAGGA 152	
QY 1235 ATCCCTCAAACTCAATTCAGCAAAATGTGAAACGGCTTAACGGTTCTTGAAGGAGAAATA 1294	
Db 151 ATCCCTCAAACTCAATTCAGCAAAATGTGAAACGGCTTAACGGTTCTTGAAGGAGAAATA 92	
QY 1295 TTTTCAAGAGGTGGAGTGTGTGTCACAGCATCTCCAAAAGCCCAATGAAGAGGCACA 1354	
Db 91 TTTTCAAGAGGTGGAGTGTGTGTCACAGCATCTCCAAAAGCCCAATGAAGAGGCACA 32	
QY 1355 GAGAAAGCATGAAATTAACAAAGGGCTCTCCC 1385	
Db 31 AAAAAAAAAAAAAAAAAACAGCGGCGAGCTCCC 1	
RESULT 13	
BY752002	
LOCUS BY752002 RIKEN full-length enriched, activated spleen Mus musculus	
DEFINITION cDNA clone F830206C17 5', mRNA sequence.	
ACCESSION BY752002	
VERSION BY752002.1	
KEYWORDS EST.	
SOURCE Mus musculus (house mouse)	
ORGANISM Mus musculus	
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muridea; Muridae; Murinae; Mus.	
AUTHORS 1 (bases 1 to 663)	
Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S., Nikaide, I., Osato, N., Saito, R., Suzuki, H., Yamanaka, I., Kiyosawa, H., Yagi, K., Tomaru, Y., Hasegawa, Y., Nogami, A., Schonbach, C., Gojobori, T., Baldarelli, R., Hill, D.P., Bult, C., Hume, D.A., Quackenbush, J., Schriml, L.M., Kanapin, A., Matsuda, H., Batalov, S., Beisel, K.W., Blake, J.A., Bredt, D., Bruscia, V., Chocho, C., Corbani, L.E., Cousins, S., Dalla, E., Dragani, T.A., Fletcher, C.F., Forrest, A., Frazer, K.S., Gaasterland, T., Gariboldi, M., Gissi, C., Godzik, A., Gough, J., Grimmond, S., Gustincich, S., Hirokawa, N., Jackson, I.J., Jarvis, E.D., Kanai, A., Kawaji, H., Kawasawa, Y., Kedzierski, R.M., King, B.L., Konegaya, A.,	
Kurochkin, I.V., Lee, Y., Lenhard, B., Lyons, P.A., Maglott, D.R., Maitais, L., Marchionni, L., McKenzie, L., Miki, H., Nagashima, T., Numata, K., Okido, T., Pavlidis, P., Pertea, G., Pesole, G., Petrovsky, N., Pillai, R., Pontius, J.U., Qi, D., Ramchandran, S., Ravasi, T., Reed, J.C., Reid, J., Ring, B.Z., Ringwald, M., Sadelin, A., Schneider, C., Semple, C.A., Setou, M., Shimada, M., Sultana, R., Takenaka, Y., Taylor, M.S., Teasdale, R.D., Tomita, M., Verardo, R., Wagner, L., Walsstedt, C., Wang, Y., Watanabe, Y., Wells, C., Wilming, L.G., Wyszewski, A., Yanagisawa, M., Yang, I., Yang, L., Yuan, Z., Zavola, M., Zhu, Y., Zimmer, A., Carninci, P., Hayatsu, N., Hirozane-Kikawa, T., Konno, H., Nakamura, M., Sakazume, N., Sato, K., Shiraishi, T., Waki, K., Waki, J., Aizawa, K., Arakawa, T., Fukuda, S., Hashizume, W., Imotani, K., Ishii, Y., Itanaka, A., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Koyama, S., Miyazaki, A., Murata, M., Nakamura, M., Nomura, K., Numazaki, R., Ohsato, N., Saito, R., Sakazume, N., Sano, H., Sasaki, D., Sato, K., Shibata, K., Shiraki, T., Tagami, M., Takeda, Y., Waki, K., Watanabe, Y., Muramatsu, M. and Hayashizaki, Y.	
Direct Submission	
Computational Analysis	
Human Genome Sequences	
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)	
RIKEN integrated sequencing analysis (RISA) system-384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)	
Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)	
cDNA library was prepared and sequenced in Mouse Genome Encyclopedia project of Riken Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in Riken Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.	
Tissues were provided by Dr. John Todd (Dept. of Medical Genetics Wellcome Trust Centre for Molecular Mechanisms in Disease Wellcome Trust/MRC Building Addenbrookes Hospital Cambridge) whose assistance we gratefully acknowledge.	
Please visit our web site (http://genome.gsc.riken.go.jp) for further details.	
FEATURES	
source	
Location/Qualifiers	
1..663	
/organism="Mus musculus"	
/mol_type="mRNA"	
/strain="NOD"	
/db_xref="taxon:10090"	
/clone="F830206C17"	
/tissue type="activated spleen"	
/clone_lib="RIKEN full-length enriched, activated spleen"	
ORIGIN	
Query Match 18.2%; Score 575.4; DB 4; Length 663;	
Best Local Similarity 97.8%; Pred. No. 1.6e-137;	
Matches 590; Conservative 0; Mismatches 12; Indels 1; Gaps 1;	

QY 1 CGAGCGCGAGTCGTCGCGCGCGCGCGCTGCGCTGCCCGCGCAGCTTTGGCGGCGAG 60
Db |||||
QY 62 CGAGCGCGAGTCGTCGCGCGCGCGCGCTGCGCTGCCCGCGCAGCTTTGGCGGCGAG 121
Db |||||
QY 61 AGCCCGTGGCTCAGATGAGATCGGGGGGACGCTGTTGGCCCTCATAGGCGACCTGAT 120
Db |||||
QY 122 AGCCCGTGGCTCAGATGAGATCGGGGGGACGCTGTTGGCCCTCATAGGCGACCTGAT 181
QY 121 GCTGCTGCTCCTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 180
Db |||||
QY 182 GCTGCTGCTCCTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 241
QY 181 GTTGATGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 240
Db |||||
QY 242 GTTGATGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 301
QY 241 GAGCGCGCAACCCCGGTACCGACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 300
Db |||||
QY 302 GAGCGCGCAACCCCGGTACCGACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 361
QY 301 CCAATATACAGAGAGATGCAAGATGCAATATAGCTTGAATCTTTATCTTAACAAAC 360
Db |||||
QY 362 CCAATATACAGAGAGATGCAAGATGCAATATAGCTTGAATCTTTATCTTAACAAAC 421
QY 361 GAGCGGTAATCTGAGGATGACTTACTCTCAGACCATCACAACATACAGAGATGCCATG 420
Db |||||
QY 422 GAGCGGTAATCTGAGGATGACTTACTCTCAGACCATCACAACATACAGAGATGCCATG 481
QY 421 GAACCGGCAAGCAGAGAGATATGACAAATTTAGAGCAAACTGGCTTCTGTTGGGATGC 480
Db |||||
QY 482 GAACCGGCAAGCAGAGAGATATGACAAATTTAGAGCAAACTGGCTTCTGTTGGGATGC 541
QY 481 CATTCAGACTTCGTTGTTCCAGAGCAACCACTCCAGTGGGAGTAAACATGAGTACGA 540
Db |||||
QY 542 CATTCAGACTTCGTTGTTCCAGAGCAACCACTCCAGTGGGAGTAAACATGAGTACGA 600
QY 541 GGTGGAAGCAG 600
Db |||||
QY 601 GGTGGAAGCAG 660
QY 601 GCA 603
Db |||||
QY 661 ACA 663

RESULT 14
BI653246
LOCUS 603300931F1 NCI_CGAP_Mam3 Mus musculus cDNA clone IMAGE:5341615 5',
DEFINITION mRNA sequence.
ACCESSION BI653246
VERSION BI653246.1 GI:15567482
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 852)
NIH-MGC <http://mgi.nhl.nih.gov/>
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Lothar Hennighausen Ph.D., Chu-Xia Deng Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LLAM11867 row: m column: 08
High quality sequence start: 12

FEATURES
source High quality sequence step: 697.
Location/Qualifiers
1..852
/organism="Mus musculus"
/mol_type="mRNA"
/strains="129,CS,BL/6J,FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:5341615"
/issue_type="umor, gross tissue"
/dev_stages="10 months"
/lab_host="DH10"
/clone_lib="NCI CGAP Mam3"
/note="Organ: mammary; Vector: pCMV-SPORT6; Site: 1: Salt;
Site 2: NotI; Cloned unidirectionally. Primer: Oligo dT.
Library constructed by Life Technologies. Investigators
providing samples: Lothar Hennighausen/Chu-Xia Deng, NIH
Reference for transgenic model: Xu et al., Nature Genetics
22, 37-43 (1999)."
ORIGIN
Query Match 18.1%; Score 573; DB 2; Length 852;
Best Local Similarity 91.3%; Prd. No. 7.4e-137;
Matches 742; Conservative 0; Mismatches 55; Indels 16; Gaps 12;
QY 1495 TTGGATAAGTAAATGAGATAA TTTTC-AAATCATCATAGAAATTTGATTTCAGCCAGG 1553
Db |||||
QY 1554 TCTCTCAGAAATGCTTCTTCTTCTTATCATGATAGCAATCCACCTTTATCAGAGTGG 1613
Db |||||
QY 1614 TAATGAAGTGTG-CAATGTG CAAAGACCTTTCTGAAGAGATGTCTGAATCATGCG 1672
Db |||||
QY 1673 CCGAGTTTTCACACAGCTCT CCTTTATAATAAATCCCTCCCATCTCCCTCTAGT 1732
Db |||||
QY 1733 AGAGTACAGAAACAAATACCC TGATGATTCAGGAAGAAAGTCTTTTACTTAGCAA 1792
Db |||||
QY 1793 TGTCCTGCTTCTGATTCAGTT GCTGTGACATTAAGCTGGGTTGGGTTTGGTTGA 1852
Db |||||
QY 1853 TTTGGGGGTTTCTTCACTTCT TTGCTATATTTTCTTACCTTTATCAGTTTGTATTC 1912
Db |||||
QY 1913 GAGCTTCTGCTTCTGGGATTC CAAATTCCTCTCCACTGACAGGATCAACTCAATGAC 1972
Db |||||
QY 1973 ATAAAGTAGTTC-AAACATCA TGTCTCACATGTTTTATFCCATAAAGTTACTCATCT 2031
Db |||||
QY 2032 GATTTTAAATTAAGTAGTGAAC TCTACTTATATCAGCCCGAGGACCATCTCCATT 2090
Db |||||
QY 2091 GGAGAAATGAAGTATTTGTCA TGGCAAGAAAGCAGGTGTGTGCATTAA--ATTGATAA 2148
Db |||||
QY 2149 GATACC-ACAAGCATCATCATG CAGTTATGAACACATGCTGTGAAA--GGATCATAGAC-- 2204
Db |||||
QY 2205 AGGGTGGTTAAATCTGATCCC GTAGATAAATCTTCACTTATTTTCAGGGAAGAG 2264

```
Db 715 GGGGTCGGTTAAATCTGTGTCAGTAGAATAACTTCAGGGTACCTATTTTCAAGGAAGAA 774
Qy 2265 TTAATTTACAACTAAAGTAAATGAACCA 2297
Db 775 TTACAGTCCCAATTAACAAATAGTAAATGACCCA 807

RESULT 15
BI100002
LOCUS 602884742F1 NCI_CGAP_Kid14 Mus musculus cDNA clone IMAGE:5039963
DEFINITION 5', mRNA sequence.
ACCESSION BI100002
VERSION BI100002.1 GI:14550895
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 771)
NIH-MGC http://mgc.ncl.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: c9apbs-remail.nih.gov
Tissue Procurement: Jeffrey E. Green, M.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM1110 row: d column: 12
High quality sequence stop: 645.
FEATURES
    source
        1..771
            /organism="Mus musculus"
            /mol_type="mRNA"
            /strain="FVB/N"
            /db_xref="taxon:10090"
            /clone="IMAGE:5039963"
            /lab_host="DH10B (T1 phage-resistant)"
            /clone_lib="NCI_CGAP_Kid14"
            /note="Organ: kidney; Vector: pCMV-SPORT6; Site 1: NotI;
            Site 2: SalI; Cloned unidirectionally. Primer: Oligo dt.
            Average insert size 1.75 kb. Constructed by Life
            Technologies. Note: this is a NCI_CGAP Library. |"
ORIGIN
Query Match 18.0%; Score 570; DB 2; Length 771;
Best Local Similarity 95.5%; Pred. No. 4.3e-136;
Matches 726; Conservative 0; Mismatches 20; Indels 14; Gaps 13;
Qy 480 CCATTCAAGACTTCGTGGTTTCCAGAGCAACACCTCCAGTGGGGACTAACATGAGCTACG 539
Db 1 CCATTCAAGACTTCGTGGTTTCCAGAGCAACACCTCCAGTGGGGACTAACATGAGCTACG 60
Qy 540 AGGTGGGAAGCAAGAAACACATCCCATTCGAGAGACATTTTCCACA-TGTTTCCAGTG 598
Db 61 AGGTGGGAAGCAAGAAACACATCCCATTCGAGAGACATTTTCCACAAGTGTTCAGTG 120
Qy 599 TCGCAGCTTTTGTGGACTATCCCTATAACCAAGTGTGAGTGGTTGGTAATGGGGGAATT 658
Db 121 TCGCAGCC-TTTGTGAGCTATCCCTATAACCAAGTGTGAGTGGTTGGTAATGGGGGAATT 179
Qy 659 CTCACCAAGTCTCTCTGGGAGCAGAAATTGATAAATCTGACTTCTGCTTCAAGGTGTAA 718
Db 180 CTCACCAAGTCTCTCTGGGAGCAGAA-TGATAAATCTGACTTCTGCTTCAAGGTGTAA 238
Qy 719 CTCCCCCCAATCACAGGGAGCGCTAGTAAGATGTGGAGCAAAACAATCTTGTGACT 778
```

```
Db 239 CTCCCCCCAATCACAGGGAGCG-TAGTAAGATG-TGGAAGCAAAACAATCTTGTGACT 297
Qy 779 GTCATATCCAGCATTTATAACCC-BAAGTACCAGAAATTTGAAGGAGAGAAAGACACAGTTT 838
Db 298 GTCATATCCAGCATTTATAACCC-BAAGTACCAGAA-TTGAAGGAGAGAAAGACACAGTTT 356
Qy 839 TTGGAGGACATCTCCACCTATG-AGATGCATTTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 898
Db 357 TTGGAGGACATCTCCACCTATG-AGATGCATTTCTCTCTCTCTCTCTCTCTCTCTCTCT 416
Qy 899 GCCAACACAGGCACTCTCTTTTA-AGTCTTACCAACACACTCAAAAGAGTCAAAAATGAGGCAA 958
Db 417 GCCAACACAGGCACTCTCTTTTA-AGTCTTACCAACACACTCAAAAGAGTCAAAAATGAGGCAA 476
Qy 959 AAGGTTCTCTTTCTTCCATCCCA-CTACCTGAGACACCTCGCTCTTTTCTTGGAGAACT-AA 1017
Db 477 AAGGTTCTCTTTCTTCCATCCCA-CTACCTGAGACACCTCGCTCTTTTCTTGGAGAACTAAA 536
Qy 1018 AGGGGTGACTGCAATACCGCTTG-CCACAGGCTTGCATGATTCGAAGTGTGGTGTGGAAC 1077
Db 537 AGGGGTGACTGCAATACCGCTTG-CCACAGGCTTGCATGATTCGAAGTGTGGTGTGGAAC 596
Qy 1078 GTGTGAAAACGTGAAGCTCTAC-CAATCTGGCCTTTCTCTAAGACTATCGAAGACACCC 1137
Db 597 GTGTG-AAAAGTGAAGCTCTA-CAATCTGGCCTTTCTCTAAGA-TATCGAAGAGCGC- 652
Qy 1138 ACTCAGTCAACCACTACTATGAT-ACATGTTACCTAAGCATGGTTTCCACCGAGATGCCTAA 1197
Db 653 -ATCAGTCAACCA-TATTATGAT-CAGTGTACTAAGCATGGGTTTCCACCGAGATGCCT-A 709
Qy 1198 AGAATACAGCCCAATGCTCCAG-TCCATATGAGAGGAAATC 1237
Db 710 AGAATACAGCCCAATGATGGAG-TCCATATGAGAGGAAATC 748

Search completed: May 31, 2006, 19:44:37
Job time : 14325 secs
```

GenCore version 5.1.1.8
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 31, 2006, 10:17:18 ; Search time 1881 Seconds
(without alignments)
11735.321 Million cell updates/sec

Title: US-10-501-930-2

Perfect score: 3166
Sequence: 1 cggagcggcagtcggtgcc.....gctattgcaaaaaaaaaa 3166

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 5244920 seqs, 3486124231 residues

Total number of hits satisfying chosen parameters: 10489840

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N_Geneseq_8:

- 1: Geneseq1980s.*
- 2: Geneseq1990s.*
- 3: Geneseq2000s.*
- 4: Geneseq2001as.*
- 5: Geneseq2001bs.*
- 6: Geneseq2002as.*
- 7: Geneseq2002bs.*
- 8: Geneseq2003as.*
- 9: Geneseq2003bs.*
- 10: Geneseq2003cs.*
- 11: Geneseq2003ds.*
- 12: Geneseq2004as.*
- 13: Geneseq2004bs.*
- 14: Geneseq2005s.*
- 15: Geneseq2006s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3166	100.0	3166	10	ADCT3012 Murine ST
2	930.2	29.4	1500	10	ADCT3014 Human ST8
3	744.8	23.5	2008	4	AAH77592 Human sia
4	614.8	19.4	840	13	ADV69200 Human cel
5	224.4	7.1	1071	6	ABV78252 Human GD3
6	224.4	7.1	1071	6	ABZ35828 Human GD3
7	224.4	7.1	1071	6	ABX10071 Human GD3
8	224.4	7.1	1071	6	ABL91793 Human pol
9	224.4	7.1	2117	2	AQ77831 Human alp
10	224.4	7.1	2117	13	ADR24998 Breast ca
11	222.2	7.0	1704	2	AAT03694 Human alp
12	207	6.5	1223	10	ADB58122 Toxicity-
13	175.4	5.5	1511	4	AAH34115 Human col
14	131.4	4.2	448	4	ABA56318 Human foe
15	131.4	4.2	448	4	ABA45804 Human bre
16	131.4	4.2	448	4	AAK29997 Human bon
17	131.4	4.2	448	6	ABSO4570 Human gen
18	130.8	4.1	428	4	ABA68901 Human foe

19	130.8	4.1	428	4	ABA50923 Human bre
20	130.8	4.1	428	4	ABA35853 Probe #14
21	130.8	4.1	428	4	AAK43020 Human bon
22	130.8	4.1	428	4	AAK17233 Human bra
23	130.8	4.1	428	6	ABS37087 Human gen
24	112.2	3.5	150	6	ABQ39173 Human ORF
25	81.8	2.6	1426	10	ADB59119 Toxicity-
26	79.6	2.5	1048	2	AAT36664 Extracell
27	79.6	2.5	1660	2	AAT36663 Sia alpha
28	79.6	2.5	1660	6	AB199514 Mouse isc
29	79.4	2.5	293	6	ABQ39108 Human ORF
30	75	2.4	472	9	ACH31114 Human foe
31	58	1.8	2000	8	ADA1938 Rice gene
32	53.6	1.7	2131	14	ADZ71009 Human chr
33	53.2	1.7	494	9	ACH13893 Human adu
34	52.6	1.7	542	12	ACH80371 Human gen
35	52.6	1.7	601	6	ABQ37664 Oligonuc1
36	52.6	1.7	601	6	ABQ37665 Oligonuc1
37	52.2	1.6	15416	6	ABL34231 Human imm
38	52.2	1.6	15416	6	ABL34231 Signal tr
39	52.2	1.6	15416	6	ABL34231 Chemical
40	52.2	1.6	15416	6	ABL34231 Human gen
41	51.6	1.6	1993	14	ADY15753 DNA encod
42	51.2	1.6	1894	14	AE832858 Breaast ca
43	51.2	1.6	10865	13	AD889721 Oligonuc1
44	50.6	1.6	581	12	ACH68308 Human gen
45	50.6	1.6	1288	9	ACF45972 Human car

ALIGNMENTS

RESULT 1

ID	ADCT3012	standard; cDNA; 3166 bp
XX	ADCT3012	
AC	ADCT3012	
DT	01-JAN-2004	(first entry)
XX	Murine ST8Sia VI cDNA.	
DE		
XX		
KW	O-glycan alpha2,8-sialyltransferase;	
KW	beta-galactoside alpha2,6-sialyltransferase; cytotostatic; virucide;	
KW	antiinflammatory; neuroprotective; cancer metastasis; viral infection;	
KW	inflammation; nerve tissue; murine; mouse; ST8Sia VI; gene; ss.	
XX		
OS	Mus sp.	
XX		
FH	Key	Location/Qualifiers
CDS		77..1273
FT		/*tag= a
FT		/product= "Murine ST8Sia VI protein"
XX		
FN	WO2003064655-A1.	
XX		
PD	07-AUG-2003.	
XX		
PF	30-JAN-2003; 2003WO-JP000883.	
XX		
PR	30-JAN-2002; 2002JP-00021159.	
PR	24-APR-2002; 2002JP-00122673.	
XX		
PA	(RIKE) RIKEN KK.	
XX		
PI	Takashima S, Tsujimoto M, Tsujimoto S;	
XX		
DR	WPI; 2003-627613/59.	
DR	P-PSDB; ADCT3011.	
XX		
PT	Sugar-chain synthases which are sialyltransferases and encoded genes,	
PT	applicable in drugs for inhibiting cancer metastasis, preventing viral	
PT	infection, inhibiting inflammation and potentiating nerve tissues.	

XX PS Claim 2; SEQ ID NO 2; 97pp; Japanese.

CC The invention relates to a novel O-glycan alpha2,8-sialyltransferase

CC having a novel substrate specificity and selectivity and a novel beta-

CC galactoside alpha2,6-sialyltransferase having a novel substrate

CC specificity and selectivity. The enzymes of the invention demonstrate

CC cytosolic, virucide, anti-inflammatory and neuroprotective activities and

CC may be applicable in drugs for inhibiting cancer metastasis, preventing

CC viral infection, inhibiting inflammation and potentiating nerve tissues.

CC The current sequence is that of the murine ST8Sia VI cDNA of the

CC invention.

XX SQ Sequence 3166 BP; 937 A; 706 C; 645 G; 878 T; 0 U; 0 Other;

Query Match 100.0%; Score 3166; DB 10; Length 3166;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 3166; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CGGAGCGGCGAGTGGTGGCGCGCGCGCGCGCTTGGCGCGCGAGG 60

Db 1 CGGAGCGGCGAGTGGTGGCGCGCGCGCGCTTGGCGCGCGAGG 60

Qy 61 ACGCCCGTGGCTCAGGATGAGATCGGGGCGACGCTGTTCCCTCATAGCAGCGCTGAT 120

Db 61 ACGCCCGTGGCTCAGGATGAGATCGGGGCGACGCTGTTCCCTCATAGGAGCGCTGAT 120

Qy 121 GCTGTGCTCTCTCTGCGTATGCTGTGTGCGCGAGCGCGCTGCGCGCTCCAGGCT 180

Db 121 GCTGTGCTCTCTCTGCGTATGCTGTGTGCGCGAGCGCGCTGCGCGCTCCAGGCT 180

Qy 181 GTTGATGAGGAGAGCAGAGGACACAGTGTGTACCTCAGCTGACATGAAGACACTCTG 240

Db 181 GTTGATGAGGAGAGCAGAGGAGCACAGTGTGTACCTCAGCTGACATGAAGACACTCTG 240

Qy 241 GAGCCCGAACCCCGGTACCGACCGACGAGGACATATCTGGATGAGAGACAAAC 300

Db 241 GAGCCCGAACCCCGGTACCGACCGACGAGGACATATCTGGATGAGAGACAAAC 300

Qy 301 CCAATACAGAGAAATGCAAGATCTGCAATATAGCTTGAATCTTATCTAACAAAC 360

Db 301 CCAATACAGAGAAATGCAAGATCTGCAATATAGCTTGAATCTTATCTAACAAAC 360

Qy 361 GAGCGGTACTCTGAGGATGACTACCTCAGACCAATCAAAACATACAGAGATGCCATG 420

Db 361 GAGCGGTACTCTGAGGATGACTACCTCAGACCAATCAAAACATACAGAGATGCCATG 420

Qy 421 GAAACCGGACGAGAGAAATGCAATTTTAGAGCAAACTGGCTTCTGTGGATGC 480

Db 421 GAAACCGGACGAGAGAAATGCAATTTTAGAGCAAACTGGCTTCTGTGGATGC 480

Qy 481 CATTCAGACTTCTGTTGTTTCCAGAACCACTCCAGTGGGACTAACATGAGTACGA 540

Db 481 CATTCAGACTTCTGTTGTTTCCAGAACCACTCCAGTGGGACTAACATGAGTACGA 540

Qy 541 GGTGGAAGAGAAACATCCCATTCGAGAGAACATTTCCATGTTTCCAGTGC 600

Db 541 GGTGGAAGAGAAACATCCCATTCGAGAGAACATTTCCATGTTTCCAGTGC 600

Qy 601 GCAGCTTTTGGGACTATCCCTATAACAGGTGTCAGTGTGGTAAATGGGGAATTC 660

Db 601 GCAGCTTTTGGGACTATCCCTATAACAGGTGTCAGTGTGGTAAATGGGGAATTC 660

Qy 661 CAAAGAGTCTCTGCGGAGCAGAAATTCATAAATCTGACTTCTGCTTCCAGTGTAACT 720

Db 661 CAAAGAGTCTCTGCGGAGCAGAAATTCATAAATCTGACTTCTGCTTCCAGTGTAACT 720

Qy 721 CCCCCCAATCAAGGAGCGGTAGTAAAGATTTGGAAAGCAAAACAAATCTTGTGACTGT 780

Db 721 CCCCCCAATCAAGGAGCGGTAGTAAAGATTTGGAAAGCAAAACAAATCTTGTGACTGT 780

Qy 781 CAATCCGAGCATTAACCCCTGAGTACAGAAATTTGAGGAGAGAGACAGTCTTTT 840

Db 781 CAATCCGAGCATTAACCCCTGAGTACAGAAATTTGAGGAGAGAGACAGTCTTTT 840

781 CAATCCGAGCATTAACCCCTGAGTACCAAAATTTGAGGAGAGAGACAGTCTTTT 840

841 GGAGGACATCTCCACCTATGAGATGCAATTCCTCTCTGCGCAGCATTTTCTATCGGCG 900

841 GGAGGACATCTCCACCTATGAGATGCAATTCCTCTCTGCGCAGCATTTTCTATCGGCG 900

901 CAACACAGGATCTCTTTTAAATCTACCAAACTCAAGAGTCAAAATAGGAGCAAAA 960

901 CAACACAGGATCTCTTTTAAATCTACCAAACTCAAGAGTCAAAATAGGAGCAAAA 960

961 GGTTCCTCTTCTTCCATCCAGGACCTGAGACACTCGCTCTTTTCTGAGAACTAAAGG 1020

961 GGTTCCTCTTCTTCCATCCAGGACCTGAGACACTCGCTCTTTTCTGAGAACTAAAGG 1020

1021 GGTGACTGACATACCGCTTGTCCAGGCTTGATGATGCAAGTGTGCTGTGGAATCTGTG 1080

1021 GGTGACTGACATACCGCTTGTCCAGGCTTGATGATGCAAGTGTGCTGTGGAATCTGTG 1080

1081 TGAAGACGTGAAGCTCTACGGATCTGCGCTTCTTAAGACTATCGAAGACACCCACT 1140

1081 TGAAGACGTGAAGCTCTACGGATCTGCGCTTCTTAAGACTATCGAAGACACCCACT 1140

1141 CAGTCACCACTACTATGATTAACCTTACCTAAGCATGGTTTCCACCAAGATGCCCTAAAGA 1200

1141 CAGTCACCACTACTATGATTAACCTTACCTAAGCATGGTTTCCACCAAGATGCCCTAAAGA 1200

1201 ATACAGCAAAATGCTCAGCTCAGATGAGGAGTAATTTTTCAGGAGGTGGAGTGTGTCA 1260

1201 ATACAGCAAAATGCTCAGCTCAGATGAGGAGTAATTTTTCAGGAGGTGGAGTGTGTCA 1260

1261 TGAAGCGCTTAAAGCTTCTTACGAGGAGTAATTTTTCAGGAGGTGGAGTGTGTCA 1320

1261 TGAAGCGCTTAAAGCTTCTTACGAGGAGTAATTTTTCAGGAGGTGGAGTGTGTCA 1320

1321 CAGCATCTCCAAAAGCCAAATAGAGAGGACACAGAGAAACATGAATTTACAAAGCGCT 1380

1321 CAGCATCTCCAAAAGCCAAATAGAGAGGACACAGAGAAACATGAATTTACAAAGCGCT 1380

1381 CTCCTGCTGTAGACCAAAAGTACCGGCGCCCACTCAGCTTTGCGAGCTCCACAGTCA 1440

1381 CTCCTGCTGTGTAGACCAAAAGTACCGGCGCCCACTCAGCTTTGCGAGCTCCACAGTCA 1440

1441 CTCATCTCAGCTTCAAGCTTCTCTGAGAAATAGAGACCAAAACATCAGACTTGGAT 1500

1441 CTCATCTCAGCTTCAAGCTTCTCTGAGAAATAGAGACCAAAACATCAGACTTGGAT 1500

1501 AAGTAAATGAGATAATTTTCAATCATATAGAAATTTGAGTGTGAGCCAGGCTCTCTCA 1560

1501 AAGTAAATGAGATAATTTTCAATCATATAGAAATTTGAGTGTGAGCCAGGCTCTCTCA 1560

1561 GAATGCTTCTTGTCTTCTTATCCAGTATGCAATTTCCACCTTTATCAGAGTGGTAAAG 1620

1561 GAATGCTTCTTGTCTTCTTATCCAGTATGCAATTTCCACCTTTATCAGAGTGGTAAAG 1620

1621 ACTGTCAATTTGTGCCAAAGACCTTCTTGAAGAGAAATGCTGAAATCATGCGCCAGTGT 1680

1621 ACTGTCAATTTGTGCCAAAGACCTTCTTGAAGAGAAATGCTGAAATCATGCGCCAGTGT 1680

1681 TTACACACAGCTCTTCTTCTTATTAATAATCCTTCCCATTTCTCCCTCTAGTAGATACA 1740

1681 TTACACACAGCTCTTCTTCTTATTAATAATCCTTCCCATTTCTCCCTCTAGTAGATACA 1740

1741 GAAACAAAATACCTTGTATGATGAGGAGAAAGTCTTTTCTTCTTCTTCTTCTTCTTCT 1800

1741 GAAACAAAATACCTTGTATGATGAGGAGAAAGTCTTTTCTTCTTCTTCTTCTTCTTCT 1800

1801 CTTCTGATTCAGTTCGCTTGTGATTAAGTCTGGGTTTGGTGGATTTGGGCG 1860

1801 CTTCTGATTCAGTTCGCTTGTGATTAAGTCTGGGTTTGGTGGATTTGGGCG 1860

1861 GTTTCCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1920

1861 GTTTCCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1920

QY 1921 TGCTTTGGGATCTTCGAATTCCTCTCCACATGACAGGATCACTCAATGACATAAGTA 1980
 Db 1921 TGCTTTGGGATCTTCGAATTCCTCTCCACATGACAGGATCACTCAATGACATAAGTA 1980
 QY 1981 GTTCAAAACATCCATGCTTCTCACAATGTTTATCCATAAAGTTACTCATCTGATTTTATT 2040
 Db 1981 GTTCAAAACATCCATGCTTCTCACAATGTTTATCCATAAAGTTACTCATCTGATTTTATT 2040
 QY 2041 TAAATAGTGAACATCTACTTGTATATCAGCCGAGGACCATCTCCATTTGGAGAATATG 2100
 Db 2041 TAAATAGTGAACATCTACTTGTATATCAGCCGAGGACCATCTCCATTTGGAGAATATG 2100
 QY 2101 AAGATATTGTCACTGCGAGAAAGAGGTGCTGCCATTAAATTGATTAAGATACCAACAGC 2160
 Db 2101 AAGATATTGTCACTGCGAGAAAGAGGTGCTGCCATTAAATTGATTAAGATACCAACAGC 2160
 QY 2161 ATCATCATGCCAGTTTATGAACACAGTGTCTGAAAGGATCATAGACAGGGGTGTTAAATCT 2220
 Db 2161 ATCATCATGCCAGTTTATGAACACAGTGTCTGAAAGGATCATAGACAGGGGTGTTAAATCT 2220
 QY 2221 GATCCAGTAGAATAAATCTCAGTGTACTATTTCAGGGAAGAGTTAATTTCACAATTA 2280
 Db 2221 GATCCAGTAGAATAAATCTCAGTGTACTATTTCAGGGAAGAGTTAATTTCACAATTA 2280
 QY 2281 AACTAGTAAATGAACCAATCTTAGCCACATTAAGTGGATTTGAGTAAAGAAAGGAA 2340
 Db 2281 AACTAGTAAATGAACCAATCTTAGCCACATTAAGTGGATTTGAGTAAAGAAAGGAA 2340
 QY 2341 CAGCAGGAGAAAGCTGTTTCGTTGTTCTGTATTACCCAAATGAGCATGCTGGAAGGAGT 2400
 Db 2341 CAGCAGGAGAAAGCTGTTTCGTTGTTCTGTATTACCCAAATGAGCATGCTGGAAGGAGT 2400
 QY 2401 TGTGAGGTACGCTAAAACCTTCGCTAGGGAGAGTACAGTGCATGAGTGTGGCGCT 2460
 Db 2401 TGTGAGGTACGCTAAAACCTTCGCTAGGGAGAGTACAGTGCATGAGTGTGGCGCT 2460
 QY 2461 TTTGTCCACATCTGTAAGGGTGAATTAATGAGCCCAATCAATCAACAGATGACAC 2520
 Db 2461 TTTGTCCACATCTGTAAGGGTGAATTAATGAGCCCAATCAATCAACAGATGACAC 2520
 QY 2521 ACCTAACTCATCACTTCAGGGGAGATGAATCTTTTCATGAGAAATTAACATCAATGCT 2580
 Db 2521 ACCTAACTCATCACTTCAGGGGAGATGAATCTTTTCATGAGAAATTAACATCAATGCT 2580
 QY 2581 AAGCATCAGTTTGTAGTAAAAATTTGAGTAGATGTTTAAATATGAACATTTTACCTCTTA 2640
 Db 2581 AAGCATCAGTTTGTAGTAAAAATTTGAGTAGATGTTTAAATATGAACATTTTACCTCTTA 2640
 QY 2641 CTAATGTCACCGACACCTTTTAAATGTAAGACATTTTATTAAGTACTTGCATTT 2700
 Db 2641 CTAATGTCACCGACACCTTTTAAATGTAAGACATTTTATTAAGTACTTGCATTT 2700
 QY 2701 AAATGCTTATGCTGTATATCTGTTTCATCCATGATTTTCCCAAAAAGTAAAGACATAG 2760
 Db 2701 AAATGCTTATGCTGTATATCTGTTTCATCCATGATTTTCCCAAAAAGTAAAGACATAG 2760
 QY 2761 GAGATGAGGCTTACATGCAAGAAACTATAAATTTTACTCTTTAAATTTCTTACTTGAGCC 2820
 Db 2761 GAGATGAGGCTTACATGCAAGAAACTATAAATTTTACTCTTTAAATTTCTTACTTGAGCC 2820
 QY 2821 AGCTTGTGTTTATCAAGTGTCTTTTGAAGAGACAGACCTGTGAATTTCTTACTTCTG 2880
 Db 2821 AGCTTGTGTTTATCAAGTGTCTTTTGAAGAGACAGACCTGTGAATTTCTTACTTCTG 2880
 QY 2881 ATACAGTGTCACTGTATTAACATTTGATGTTTCAAGTTTACATCTCTTTTCAAT 2940
 Db 2881 ATACAGTGTCACTGTATTAACATTTGATGTTTCAAGTTTACATCTCTTTTCAAT 2940
 QY 2941 TCTTTTATAGCAAAATCAACAGTATTAGCTTTCAGAAATTTATCAGAAAGTTTCAATATAAT 3000
 Db 2941 TCTTTTATAGCAAAATCAACAGTATTAGCTTTCAGAAATTTATCAGAAAGTTTCAATATAAT 3000

QY 3001 ATTTTGCAGAGGTAAAGGCTTTTGTAAATAAATAAATAATTTATTTCTTCTG 3060
 Db 3001 ATTTTGCAGAGGTAAAGGCTTTTGTAAATAAATAAATAATTTATTTCTTCTG 3060
 QY 3061 ATGAATAGAGGCTCTTTTATGCTGCTAATGAACCTAATTAAGTTTAAATATCTCCT 3120
 Db 3061 ATGAATAGAGGCTCTTTTATGCTGCTAATGAACCTAATTAAGTTTAAATATCTCCT 3120
 QY 3121 AGCAACATTTGCTCAGTTTCAATCATGCTATTAGCAAAAAA 3166
 Db 3121 AGCAACATTTGCTCAGTTTCAATCATGCTATTAGCAAAAAA 3166

RESULT 2
 ADC73014
 ID ADC73014 standard; cDNA; 1500 B.
 XX
 AC ADC73014;
 XX
 DT 01-JAN-2004 (first entry)
 XX
 DE Human ST8Sia VI cDNA.
 XX
 KW O-glycan alpha2,8-sialyltransferase;
 KW beta-galactoside alpha2,6-sialyltransferase; cytostatic; virucide;
 KW antiinflammatory; neuroprotective; cancer metastasis; viral infection;
 KW inflammation; nerve tissue; human; ST8Sia VI; gene; ss.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT CDS 92..1288
 FT /*tag= a
 FT /product= "Human ST8Sia VI protein"
 XX
 PN WO2003064655-A1.
 XX
 PD 07-AUG-2003.
 XX
 PF 30-JAN-2003; 2003WO-JP000883.
 XX
 PR 30-JAN-2002; 2002JP-00021159.
 PR 24-APR-2002; 2002JP-00122673.
 XX
 PA (RIKE) RIKEN KK.
 XX
 PI Takashima S, Tsujimoto M, Tsuji S;
 XX
 DR WPI; 2003-627613/59.
 DR P-PSDB; ADC73013.
 XX
 PT Sugar-chain synthases which are sialyltransferases and encoded genes,
 PT applicable in drugs for inhibiting cancer metastasis, preventing viral
 PT infection, inhibiting inflammation and potentiating nerve tissues.
 XX
 PS Claim 4; SEQ ID NO 4; 97bp; Japanese.
 XX
 CC The invention relates to a novel O-glycan alpha2,8-sialyltransferase
 CC having a novel substrate specificity and selectivity and a novel beta-
 CC galactoside alpha2,6-sialyltransferase having a novel substrate
 CC specificity and selectivity. The enzymes of the invention demonstrate
 CC cytostatic, virucide, antiinflammatory and neuroprotective activities and
 CC may be applicable in drugs for inhibiting cancer metastasis, preventing
 CC viral infection, inhibiting inflammation and potentiating nerve tissues.
 CC The current sequence is that of the human ST8Sia VI cDNA of the
 CC invention.
 XX
 SQ Sequence 1500 BP; 425 A; 357 C; 349 G; 369 T; 0 U; 0 Other;
 Best Local Similarity 29.4%; Score 930.2; DB 10; Length 1500;
 Matches 1169; Conservative 0; Mismatches 308; Indels 14; Gaps 4;


```
FT      /*tag= a
XX      /product= "cell division regulating protein 27"
PN      CN1493590-A.
PD      05-MAY-2004.
XX
XX      30-OCT-2002; 2002CN-00137726.
XX
XX      30-OCT-2002; 2002CN-00137726.
XX      (SHAN-) SHANGHAI BIODOOR GENE DEV CO LTD.
XX
XX      Mao Y, Xie Y;
XX
XX      WPI; 2004-500615/48.
XX      P-PSDB; ADV69201.
XX
XX      New polypeptide-human cell disintegrate regulatory protein 27 for
XX      treating cancer, human immunodeficiency virus infection, and immunopathy.
XX
XX      Claim 6; SEQ ID NO 1; 25pp; Chinese.
XX
XX      The invention relates to human cell division regulating protein 27, the
XX      polynucleotide encoding it, preparing the polypeptide by DNA recombinant
XX      techniques, application of the polypeptide in treating diseases, such as
XX      cancer, human immunodeficiency virus (HIV) infection, immunopathy and the
XX      antagonist of the polypeptide and its medical action and application of
XX      the polynucleotide. The present sequence is that of the human cell
XX      division regulating protein 27 encoding DNA of the invention.
XX
XX      Sequence 840 BP; 267 A; 178 C; 164 G; 231 T; 0 U; 0 Other;
XX
XX      Query Match      19.4%; Score 614.8; DB 13; Length 840;
XX      Best Local Similarity 83.6%; Pred. No. 1.3e-150;
XX      Matches 697; Conservative 0; Mismatches 137; Indels 0; Gaps 0;
XX
XX      440 TATGACAAATTTAGAGCAAACTGGCTTCCTGTTGCGATGCCATTCAGACTTGTGTT 499
XX      1 TATGCAATTTTAGAGCCAAACTTGTCTCTGCTGTGATGCTGTCAAAACTTTGTGTT 60
XX
XX      500 TCCAGACAACTCCAGTGGGACTAACTAGCTAGCTAGAGGTGGAAGCAAGAACAC 559
XX      61 TCTCAGATAACACTCCAGTTGGGACTAATATGATTACGAGGTGGAAGCAAGAACAA 120
XX
XX      560 ATCCCCATTCGAGAGAACATTTTCCACATGTTTCCAGTGTCCGACCTTTTGTGACTAT 619
XX      121 ATCCCAATTAAGAGAACATTTTTCATATGTTTCCAGTGTCCAGCTTTTGTGACTAC 180
XX
XX      620 CCCTATAACAGTGTGCTGCTGTTGTTGTAATGGGGAAATTCACAAAGTCTCTCGGGA 679
XX      181 CTTTATATATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 240
XX
XX      680 GCAGAAATTAATATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 739
XX      241 ACTGAAATAGATAAATCCGACTTCTGTTTATGTTAATCCCTACCCCAACCAAGAGAT 300
XX
XX      740 GCTAGTAAAGATGTTGGAAGCAAAACAATCTGTGACTGTCAATCCAGCATTTATAACC 799
XX      301 GTTAGTAAAGATGTTGGAAGCAAAACAATCTGTGACTATAATCCCAAGCATATAACT 360
XX
XX      800 CTGAAGTACCAAGATTTGAAGAGAGAGAAAGACAGTATTTTGGAGGACATCTCCACCTAT 859
XX      361 CTGAATATGGAATTTGAAGAGAGAGAAAGAGCCCTATTTCTGGAGGACATTTGCAACCTAT 420
XX
XX      860 GGAGATGATTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 919
XX      421 GGAGATGATTTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 480
XX
XX      920 AAAGTCTACCAAACTCAAAAGAGTCAAAAGATGAGGCAAAAGGTTCTCTCTTCCATCCC 979
XX      481 AAAGTATATACAGCTCGAGAGTCTAAGACAGACAAAGGTTCTATTTTCCATCCC 540
```

```
QY      980 AGGTACTCTGAGACACCTCGCTCTTTCTGGAGAACTAAAGGTGACTGCATACCGCTTG 1039
DB      541 AAGTACCTCTGAAAGATCTGGGCTTTTCTGGAGAACTAAAGGTGACTGCATACCGCTTG 600
QY      1040 TCCACAGGCTTGATGATTGCAATGTCGCTGTGTGAACTGTGTGAAACGCTGAAGCTCTAC 1099
DB      601 TCCACCGGCTTGATGATCACAAGTGTTCAGTGCAACTGTGTGAAATGTGAAGCTGTAT 660
QY      1100 GGATTTGGGCTTTCTCTAAGATATCGAAGACACCCCATCTCAGTCACCATCTATGAT 1159
DB      661 GGATTTGGGCTTTCTCTAAGATATCGAAGACACCATCTCAGTCACCATCTATGAT 720
QY      1160 AACATGTTACTAAGCATGTTTCCACAGATGCTAAAGATACAGCAATGCTCCAG 1219
DB      721 AACAGCTACCTAAACATGTTTCCATCAGATGCCAAGAAATACAGCCAGATCTCCAA 780
QY      1220 CTCATATGAGAGGAATCCTCAAACTGCAATTTAGCAAAATGTGAAGTCGCTTAA 1273
DB      781 CTTACATGAAGGAATCCTCAAACTGCAATTTAGCAAAATGTGAAGTCGCTTAA 834
XX
XX      RESULT 5
XX      ABV78252
XX      ID ABV78252 standard; DNA; 1071 BP
XX      AC ABV78252;
XX      DT 15-NOV-2002 (first entry)
XX      DE Human GD3 synthase DNA SEQ ID NO 136.
XX      KW RNA inhibition; dsRNA; gene expression inhibitor; oncogene; cytostatic;
XX      OS virucide; protozoicide; Gene; da
XX      PN Homo sapiens.
XX      WO2002555693-A2.
XX      PD 18-JUL-2002.
XX      PP 09-JAN-2002; 2002WO-EP000152.
XX      PR 09-JAN-2001; 2001DE-01000586.
XX      PR 26-OCT-2001; 2001DE-0105280.
XX      PR 29-NOV-2001; 2001DE-01058411.
XX      PR 07-DEC-2001; 2001DE-01060151.
XX      PA (RIBO-) RIBOPHARMA AG.
XX      PI Kreutzer R, Limmer S, Rost S, Hadwiger P;
XX      WIPI; 2002-590671/63.
XX      PT Inhibiting expression of target gene, useful e.g. for inhibiting
XX      oncogenes, by administering double-stranded RNA complementary to the
XX      target and having an overhang.
XX      PS Claim 10; Page 193-194; 203pp; German.
XX      CC The invention relates to inhibiting expression of a target gene (I) in a
XX      cell by introducing an inhibitory RNA (dsRNA) having a double-stranded
XX      structure of at most 49 consecutive bases. At least part of one strand
XX      (asi) of dsRNA is complementary to (I) and at least one end of dsRNA
XX      has an overhang of 1-4 nucleotides. The method is used to inhibit the
XX      expression of a wide range of genes, e.g. oncogenes, cytokine genes etc.
XX      in humans, also genes in Plasmodium or in viruses or viroids that are
XX      pathogenic for humans, animals or plants. Introducing an overhang into
XX      dsRNA greatly increases effectiveness for inhibiting gene expression,
XX      both in vivo and in vitro and also increases stability and thus the
XX      effective concentration inside the cell. The present sequence is that of
XX      a gene related to the invention
XX      Sequence 1071 BP; 264 A; 277 C; 269 G; 261 T; 0 U; 0 Other;
```

Query Match 7.1%; Score 224.4; DB 6; Length 1071;
 Best Local Similarity 55.3%; Prd. No. 6.4e-48;
 Matches 457; Conservative 0; Mismatches 366; Indels 3; Gaps 1;

QY 449 TTTAGAGCAAACTGGCTTCTGTTGCGATGCCATTCAAGACTTCTGTTGTTTCCAGAAC 508
 DB 232 TTCAGGAACAATGGAAGACTGCTGGACCTGCGCCATCTCTTGTCTATGACTAAATG 291
 QY 509 AACACTCCAGTGGGACTTAACATGAGTACGAGGTGGAAAGCAAGAAACACATCCCATTT 568
 DB 292 AATTCCTCTATGGGAAGACATGTTGATGACGGGAGTCTTTTACTCAATCCACCAT 351
 QY 569 CGAGAGACATTTTCCACATGTTTCCAGTGTGCGAGCCTTTTGTGACTATCCCTATAAC 628
 DB 352 GACAAATTCACCTTACTCTCTCTTCCACAGGCAACCCCAT---TCCAGTGCCTTGAAG 408
 QY 629 CAGTGTGAGTGGTGGTAATGGGGAATTTCTCAACAGTCTCTCTGCGGAGCAGAAAT 688
 DB 409 AAATCGCGGTGGTGGAAATGGTGGATCTCGAAGAGAGTGGCTGTGGCGCTCAATA 468
 QY 689 GATAAATCTGACTTCTGTTTCAAGTGTAACTTCCCTCCCAATCACAGGAGCGCTAGTAA 748
 DB 469 GATGAAGCAAAATTTGTCAATGATGCAATCTCCCTCTTGTCAAGTGAATACACTAAG 528
 QY 749 GATGTTGGAAGCAAAACAAATCTTGTGACTGTCAATCCAGCATTTATACCTGAAGTAC 808
 DB 529 GATGTTGGAATCAAAAGTCAAGTGTAGTGAAGCAATATCCAGCAATAATTCGGCAAAAGTTT 588
 QY 809 CAGAAATTTCAAGGAGAGCAACACAGTGTGAGGAGCATCTCCACCTATGAGATGCA 868
 DB 589 CAGAACTTCTGTGTCGAGAAAGCAATTTGTGGCAACATGAAATCTATTAACACAGT 648
 QY 869 TTCCTCTCTCCAGCAATTTTCTATCGGGCAACACAGGATCTCTTTTAAAGTCTAC 928
 DB 649 TACATCTAGTCCCTGCTTTTCTATGAGACAGGACAGCCATCTTTGAGGTTTAT 708
 QY 929 CAAACACTCAAGAGTCAAAATGAGGCAAAAGTTTCTTCTTCAATCCAGGTTACCTG 988
 DB 709 TATACACTGTGAGTGTGGTGGCAATCAACAGTGTCTTTTGGCAACCCCAACTTTCTG 768
 QY 989 AGACACTCGCTCTTTCTGGAGAACTAAAGGGTCACTGCATACCGTGTGTCACAGC 1048
 DB 769 CGTAGCATGGAAGTCTGGAAAGTAGAGAAATCCATGCAAGCGCTGTCCACAGA 828
 QY 1049 TTGATGATGCAAGTGTGCTGTGAACTGTGTGAAAGCTGAAGCTCTACCGATTTCTGG 1108
 DB 829 CTTTCTGTTGAGCGCAGCTCTGGTCTCTGTGAAGGTGGCCATCTATGGCTTCTGG 888
 QY 1109 CTTTCTCTAAGACTATGGAAGACACCCCACTCAGTCACTACTATGATTAACATGTTA 1168
 DB 889 CTTTCTCTGTAATATGATGAGCAGCGCCATCAGCCACCACTACTATGACAACTCTTA 948
 QY 1169 CTAAGCATGTTTCCACAGATGCTTAAGATATACAGCCAAATGCTCCAGTCCCATATG 1228
 DB 949 CTTTCTGTTGCTTCCATGCCATGCGGAGGAATTTCTCCAACTCTGTTATCTTCAATA 1008
 QY 1229 AGAGGAATCTCAAACTGCAATTTCAAGAAATGTGAAACGGCTTAAAC 1274
 DB 1009 ATCGTGCNCTGAGATGAGCTGGNCCCATGTGAAGATACCTCAC 1054

RESULT 6
 AB235828
 ID AB235828 standard; DNA; 1071 BP.
 AC
 AC AB235828;
 DT 07-FEB-2003 (first entry)
 XX Human GD3 synthase polynucleotide SEQ ID NO 136.
 XX Double stranded RNA; dsRNA; RNAi; RNA inhibition; cytostatic; virucide;
 KW

KW protozoacide; gene expression; antisense; tumour; infection; plasmodium;
 KW virus; viroid; anti-GFP; human; HIV; human immunodeficiency virus;
 XX Hepatitis C virus; human papilloma virus; gene; ds.
 OS Homo sapiens.
 PN DE10100588-A1.
 XX 18-JUL-2002.
 PD
 PF 09-JAN-2001; 2001DE-01000588.
 XX
 PR 09-JAN-2001; 2001DE-01000588.
 XX
 PA (RIBO-) RIBOPHARMA AG.
 XX
 PI Kreutzer R, Limmer S, Rost S, Hadwiger P;
 XX WPI; 2002-683450/74.
 DR
 XX Inhibiting expression of target genes, useful e.g. for treating tumors,
 PT by introducing into cells two double-stranded RNAs that are complementary
 PT to the target.
 XX
 PS Claim 13; Page 91; 100pp; German.
 XX
 CC The invention relates to inhibiting expression of a target gene in a cell
 CC by introducing at least two oligonucleotides (dsRNAI and II), both
 CC with a double-stranded (ds) structure of at most 49 sequential nucleotide
 CC pairs. At least part of one strand (S1, S2) of the ds structures in each
 CC of dsRNAI and II are complementary to regions in the target gene. The
 CC method uses antisense inhibition of gene expression using double stranded
 CC RNA inhibition (RNAi). The method is particularly used to treat tumours
 CC or infections, especially by plasmodium or viruses/viroids (pathogenic on
 CC humans, animals or plants). The method provides more effective inhibition
 CC of expression than known methods using a single dsRNA, even at very low
 CC concentrations. When dsRNA has at least one unpaired nucleotide at the
 CC end, stability (and thus effective concentration in the cell) is improved
 CC and efficiency can be increased further by pretreating the cells with
 CC interferon. The present sequence is that of a target DNA of the invention
 XX
 SQ Sequence 1071 BP; 264 A; 277 C; 269 G; 261 T; 0 U; 0 Other;
 Query Match 7.1%; Score 224.4; DB 6; Length 1071;
 Best Local Similarity 55.3%; Prd. No. 6.4e-48;
 Matches 457; Conservative 0; Mismatches 366; Indels 3; Gaps 1;

QY 449 TTTAGAGCAAACTGGCTTCTGTTGCGATGCCATTCAAGACTTCTGTTGTTTCCAGAAC 508
 DB 232 TTCAGGAACAATGGAAGACTGCTGGACCTGCGCCATCTCTTGTCTATGACTAAATG 291
 QY 509 AACACTCCAGTGGGACTTAACATGAGTACGAGGTGGAAAGCAAGAAACACATCCCATTT 568
 DB 292 AATTCCTCTATGGGAAGACATGTTGATGACGGGAGTCTTTTACTCAATCCACCAT 351
 QY 569 CGAGAGACATTTTCCACATGTTTCCAGTGTGCGAGCCTTTTGTGACTATCCCTATAAC 628
 DB 352 GACAAATTCACCTTACTCTCTCTTCCACAGGCAACCCCAT---TCCAGTGCCTTGAAG 408
 QY 629 CAGTGTGAGTGGTGGTAAATGGGGAATTTCTCAACAGTCTCTCTGCGGAGCAGAAAT 688
 DB 409 AAATCGCGGTGGTGGAAATGGTGGATCTCGAAGAGAGTGGCTGTGGCGCTCAATA 468
 QY 689 GATAAATCTGACTTCTGTTTCAAGTGTAACTTCCCTCCCAATCACAGGAGCGCTAGTAA 748
 DB 469 GATGAAGCAAAATTTGTGATGCAATCTCCCTCTTGTCAAGTGAATACACTAAG 528
 QY 749 GATGTTGGAAGCAAAACAAATCTTGTGACTGTCAATCCAGCATTTATACCTGAAGTAC 808
 DB 529 GATGTTGGAATCAAAAGTCAAGTGTAGTGAAGCAATATCCAGCAATAATTCGGCAAAAGTTT 588
 QY 809 CAGAAATTTCAAGGAGAGCAACACAGTGTGAGGAGCATCTCCACCTATGAGATGCA 868

589 CAGAACCTTCTGTCGTCAGAGACATTTGTGGACAAATGAAATCTATTAACACAGT 648
 869 TTCCTCTCTGTCGTCAGACATTTCTATTCGGGCAACACAGGCAATCTCTTTTAAAGTCTAC 928
 649 TACATCTACATGCTGCTGCTTTCTATGAAGACAGAGAACAGACCAATCTTTGAGGGTTTAT 708
 929 CAAACACTCAAGAGTCAAAATATGAGGCAAAAGTTCTCTTCTCCATCCAGGTACCTG 988
 709 TATACACTGTCTAGATGTTGGTGCATCAACACAGTCTGTTGGCAACCCCACTTTCTG 768
 989 AGACACTCTGCTCTTTTCTGGAGAACTAAAGGGGTGACTGCATACCGCTTGTCCACAGGC 1048
 769 CGTAGCATTTGGAAGTTCTGGAAGAGTACAGGAATCCATGCCAGGCTGTCCACAGGA 828
 1049 TTGATGATTGCAAGTCTGCTGTGGAACCTGTGTGGAACCTGTAAGCTCTACGGATTCGG 1108
 829 CTTTCTCTGCTGAGCGCAGCTCTGGGTCTCTGTGAAGAGTGGCCATCTATGGCTTCTGG 888
 1109 CTTTCTCTGCTGAGCGCAGCTCTGGGTCTCTGTGAAGAGTGGCCATCTATGGCTTCTGG 888
 889 CTTTCTCTGCTGAGCGCAGCTCTGGGTCTCTGTGAAGAGTGGCCATCTATGGCTTCTGG 888
 1169 CCTAAGCATGTTTCCACAGATGCTTAAAGAAATACAGCAAAATGCTCCAGCTCCATATG 1228
 949 CTTTCTCTGCTGAGCGCAGCTCTGGGTCTCTGTGAAGAGTGGCCATCTATGGCTTCTGG 1008
 1229 AGAGAACTCTCAAACTGCAATTCAGCAAAATGTGAAACGGCTTAAC 1274
 1009 ATCGGTGCACTGAGAATGCAAGTGGAGCCATCGTGAAGATACCTCAC 1054

RESULT 7

ABX10071
 ID ABX10071 standard; DNA; 1071 BP.
 AC ABX10071;
 DT 23-JAN-2003 (first entry)
 XX Human GD3 synthase DNA fragment SEQ ID 136.
 DE Oligoribonucleotide; interferon; oncogene; cytokine; Id; developmental;
 KW prion; inhibition; human; ds.
 XX Homo sapiens.
 OS DE10100587-CL.
 XX 21-NOV-2002.
 XX 09-JAN-2001; 2001DE-01000587.
 XX 09-JAN-2001; 2001DE-01000587.
 XX (RIBO-) RIBOPHARMA AG.
 PA Kreutzer R, Limmer S, Rost S, Hadwiger P;
 FI WPI; 2002-742209/81.
 DR Inhibiting expression of target genes, e.g. oncogenes, in cells, by
 PT introduction of complementary double-stranded oligoribonucleotide, after
 PT treating the cell with interferon.
 XX Disclosure; Page 96; 98pp; German.
 PS This invention describes a novel method for inhibiting expression of a
 CC target gene by introducing into the cell that contains the target gene at
 CC least one oligoribonucleotide (dsRNA) that has a double-stranded (ds)
 CC structure of not more than 49 consecutive nucleotides (nt), where at
 CC least a segment of one strand of the ds structure is complementary with
 CC the target gene and the cells are treated with interferon before
 CC introduction of dsRNA. The method is used to inhibit expression of

CC target genes, particularly oncogenes, cytokine genes, Id (not defined)
 CC protein genes; developmental or prion genes, or genes expressed in
 CC pathogenic organisms (particularly plasmids) or in viruses or viroids
 CC (pathogenic in humans, animals or plants). Treating the cells with
 CC interferon greatly increases the extent to which dsRNA can inhibit
 CC expression of the target genes, and the effect is even greater when dsRNA
 CC are modified to increase their stability. ABX0936-ABX10075 represent
 CC gene fragments used to illustrate the method of the invention
 XX
 SQ Sequence 1071 BP; 264 A; 277 C; 269 G; 261 T; 0 U; 0 Other;
 Query Match 7.1%; Score 224.4; DB 6; Length 1071;
 Best Local Similarity 55.3%; Pruned No. 6.4e-48;
 Matches 457; Conservative 0; Mismatches 366; Indels 3; Gaps 1;
 QY 449 TTTAGAGCAAACTGGCTTCCTTTGCGATGCCATTCAGAGCTTCGTGGTTTCCAGAAC 508
 DB 232 TTCAGGAAACAAATGGAAGACTTTCGACCTCGCCCATCTCTTGTCTATGACTAAATG 291
 QY 509 AACACTCCAGTGGGACTTAACAATAGCTACAGGTGGAAAGCAAGAACACATCCCAT 568
 DB 292 AATTCCCTATATGGGGAAGACAATGGGTATGACGGGGAGTTTATATCTCATTTCCACCA 351
 QY 569 CGAGAGAACATTTTCCACATGTTCCAGTGTGCGAGCCTTTTGTGGACTATCCCTATTAAC 628
 DB 352 GACAAATTCATTTACTCTCTTCCACAGGCAACCCCAT---TCCAGTGCATTTGAAG 408
 QY 629 CAGTGTGAGTGGTGGTAAATGAGGAAATTCACAAAGTCTCTCTGCGGAGACAGAAAT 688
 DB 409 AAATGCGGGTGGTGGGAAATGAGGAAATTCGAAAGAGAGTGGCTGTGGCGCTCAATA 468
 QY 689 GATAAATCTGATTCGTTCTTCAATGTAACTCCCTCCCAATCACAGGAGCGCTAGTAA 748
 DB 469 GATGAAGCAAAATTTGTCTATGCAATCTCCCTCTCTTGTCAAGTGAATACACTAAG 528
 QY 749 GATGTTGAAGCAAAACAAATCTGTGACTGTCAATCCAGCATTAATACCTCCAGGTAC 808
 DB 529 GATGTTGAATCAAAAGTCACTAGTGACAGTAATCCAGCATTAATTCGGCAAGGTTT 588
 QY 809 CAGAAATTTGAAGGAGAAAGAACAGTGTGAGGAGCATCTCCACCTATGGAGATGA 868
 DB 589 CAGAACTTCTGTGTCCAGAAAGACATTTGTGGACACATGAAATCTATAACACAGT 648
 QY 869 TTCCTCTCTGCGCAGCATTTTCTATCGGCGCAACACAGGCAATCTCTTTAACTCTAC 928
 DB 649 TACATCTACATGCTGCTTTTATGAAGACAGGAAACAGAGCCATCTTTGAGGGTTTAT 708
 QY 929 CAAACACTCAAGAGTCAAAATGAGGCAAAAGTTCTCTTCTCCATCCAGGTACCTG 988
 DB 709 TATACACTGTCTAGATGTTGGTGCATCAACACAGTCTGTTGGCAACCCCACTTTCTG 768
 QY 989 AGACACTCTGCTCTTTTCTGGAACAACTAAAGGGTGAATGATACCTGTTGTCACAGGC 1048
 DB 769 CGTAGCATTTGGAAGTTCTGGAAGAGGAAATCCATGCCAGGCGCTGTCCACAGGA 828
 QY 1049 TTGATGATTGCAAGTGTGCTGTGGAACTGTGAAACCTGTAAGTCTACGGATTCCTGG 1108
 DB 829 CTTTCTCTGCTGAGCGCAGCTCTGGGTCTCTGTGAAGAGTGGCCATCTATGGCTTCTGG 888
 QY 1109 CTTTCTCTGCTGAGCGCAGCTCTGGGTCTCTGTGAAGAGTGGCCATCTATGGCTTCTGG 888
 DB 889 CTTTCTCTGCTGAGCGCAGCTCTGGGTCTCTGTGAAGAGTGGCCATCTATGGCTTCTGG 888
 QY 1169 CCTAAGCATGTTTCCACAGATGCTTAAAGAAATACAGCAAAATGCTCCAGCTCCATATG 1228
 DB 949 CTTTCTCTGCTGAGCGCAGCTCTGGGTCTCTGTGAAGAGTGGCCATCTATGGCTTCTGG 1008
 QY 1229 AGAGAACTCTCAAACTGCAATTCAGCAAAATGTGAAACGGCTTAAC 1274
 DB 1009 ATCGGTGCACTGAGAATGCAAGTGGAGCCATCGTGAAGATACCTCAC 1054

RESULT 8

DR WPI; 1994-333182/41.
 XX P-PSDB; AAR63225.
 PT New alpha-2,8,-sialyl transferase - useful for producing physiologically
 XX active sugar chains.
 PS Claim 3; Page 63-66; 102pp; Japanese.
 XX
 CC The nucleotide sequence of the novel gene encoding the human alpha-2,8-
 CC sialyltransferase. The gene encodes a protein of 356 a.a. The gene
 CC (called WPI) was isolated from the human melanoma cell line WM266-4 and
 CC cloned into the E.coli plasmid pUC119 to create pUC119.WPI. The gene was
 CC inserted into the expression vector pAMP/PRC3SC to produce the plasmid
 CC pAMP/PRSAWPI. The missing N-terminal 19 a.a. were inserted into this
 CC vector by PCR amplification using the primers AAQ77835-6. Alpha-2,8-
 CC sialyltransferase is useful in the production of physiologically active
 CC sugar chains e.g. in the conversion of ganglioside GM3 to GD3. (Updated
 CC on 25-MAR-2003 to correct PN field.)
 XX
 SQ Sequence 2117 BP; 554 A; 502 C; 571 G; 490 T; 0 U; 0 Other;
 Query Match 7.1%; Score 224.4; DB 2; Length 2117;
 Best Local Similarity 55.3%; Pred. No. 9.2e-48;
 Matches 457; Conservative 0; Mismatches 366; Indels 3; Gaps 1;
 QY 449 TTTAGAGCAAACTGGCTTCTGTTGGCATGCCATTCAGACTTCGTGGTTTCCAGAAC 508
 DB 714 TTCAGGAACAATGGAAGACTGCTGCGACCTGCCATCTCTTTGCTATGACTAAATG 773
 QY 509 AACACTCCAGTGGGACTTAACATGAGCTACGAGTGGAAAGCAAGAACACATCCCCATT 568
 DB 774 AATTCCCTATGGGAGAGCATGTGATGACGGGAGTTTTTATCTCATTCACCAT 833
 QY 569 CGAGAGAACATTTCCACATGTTTCCAGTGTGCGAGCCTTTTGGACTATCCCTATAAC 628
 DB 834 GACAAATTCAACTTACTCTCTCTCCACAGGCAACCCCAT---TCCAGCTGCCATTGAAG 890
 QY 629 CAGTGTGAGTGGTGTGTAATGGGGGAATTCACACAGTCTCTCTGGGAGCAAGAAAT 688
 DB 891 AATATGCGGCTGGTGGGAATGTTGGGATCTGGAAGAAGATGGCTGTGGCGGTCAATA 950
 QY 689 GATAAATCTGACTTCGTCTTCAGGTGTAACTCCGCCCAATCACAGGAGGCGTAGTAAA 748
 DB 951 GATGAGCAAAATTTGTTCATCGCATGCAATCTCCCTCTCTTGTCAAGTGAATACACTAAG 1010
 QY 749 GATGTTGAAGCAAAACAAATCTGTGTGACTGTCAATCCAGCATTAATACCTGAAGTAC 808
 DB 1011 GATGTTGGATCCAAAGTCACTAGTACAGCTAATCCAGCATTAATCCGCAAGGTTT 1070
 QY 809 CAGAAATTTGAAGGAGAGAAAGCACAGTCTTTTGGAGGACATCTCCACCTATGGAGATGCA 868
 DB 1071 CAGAACCTTCTGTGTCGAGAAAGACATTTGTGACAAATGAAATCTAATACCAAGT 1130
 QY 869 TTCCTCTCTCCGCGACATTTTCCTATCGGCGCAACACAGGACATCTCTTTTAAAGTCTAC 928
 DB 1131 TACATCTACATGCTGCTCTTCTATGAGACAGAGCAAGAGCCATCTTTGAGGCTTAT 1190
 QY 929 CAACACCTCAAGAGTCAAAATGAGGCAAAAGTTCTCTTCTTCATCCAGGTACTG 988
 DB 1191 TATACACTGTGAGTGTGGTGGCAATCAAAACAGTGTGTTTGGCAACCCCAACTTTCTG 1250
 QY 989 AGACACCTCGCTCTTTCTGAGAACTAAAGGGTGTGATCGATACCGCTTGTCCACAGGC 1048
 DB 1251 CGTAGCATTTGGAAGTTCTGGAAGGTAGAGAAATCCATGCAAGCGGCTGTCCACAGGA 1310
 QY 1049 TTGATGATGCAAGTGTGCTGTGGAACTGTGTGAAACAGTGAAGCTCTACGGATTTCTGG 1108
 DB 1311 CTTTCTTCTGAGGAGCTCTGGGTCTCTGTGAGAGGTGGCCATCTATGGCTTCTGG 1370
 QY 1109 CTTTCTCTAAGACTATCGAAGACACCCCACTAGTACCACTACTATGATPAAACATGTTA 1168
 DB 1371 CCCTTCTCTGTGTAATATGATGAGCAGGCCCATCAGCCACCACTACTATGACAACTGTTA 1430

QY 1169 CCTAAGCATGTTTCCACCAGATGTCCTAAAGAAATACAGCAAAATGCTCCAGTCCATATG 1228
 DB 1431 CCTTTTCTGGCTTCCATGCCATGCGGAGGAATTTCTCCAACTCTGTTATCTTCAATAA 1490
 QY 1229 AGAGGAATCTCAAACTGCAATTCAGCAAAATGTCAAAACGGCTTAAC 1274
 DB 1491 ATCGGTGCACTGAGAAATGCACTGAGCCCATGTGAAGATACCTCAC 1536

RESULT 10
 ADR24998
 ID ADR24998 standard; DNA; 2117 BP.
 XX
 AC ADR24998;
 XX
 DT 21-OCT-2004 (first entry)
 XX
 DE Breast cancer prognosis marker #359.
 XX
 KW ds; breast cancer; prognosis; gene expression; diagnosis.
 XX
 OS Homo sapiens.
 XX
 PN W02004065545-A2.
 XX
 PD 05-AUG-2004.
 XX
 PF 15-JAN-2004; 2004WO-US001100.
 XX
 PR 15-JAN-2003; 2003US-00342887.
 XX
 PA (ROSE-) ROSETTA INPHARMATICS LLC
 PA (NECA-) NETHERLANDS CANCER INST.
 XX
 PI Van't Veer LJ, He Y;
 XX
 DR WPI; 2004-593473/57.
 XX
 PT Classifying a breast cancer patient according to prognosis comprises
 PT determining the similarity between the level of expression of each of
 PT five genes in a cell sample taken from patient, to control levels.
 XX
 PS Disclosure; SEQ ID NO 859; 226pg English.
 XX
 CC The invention relates to a method of classifying a breast cancer patient
 CC according to prognosis by determining the similarity between the level of
 CC expression of each of five genes for which markers are listed in the
 CC specification, in a cell sample taken from the breast cancer patient, to
 CC control levels of expression for each respective five genes to obtain a
 CC patient similarity value. The methods are useful for classifying a breast
 CC cancer patient according to prognosis. Kits and computer program products
 CC are useful for data analysis using the diagnostic, prognostic and
 CC statistical methods of the invention. This sequence corresponds to a
 CC marker used in the method of the invention.
 XX
 SQ Sequence 2117 BP; 554 A; 502 C; 571 G; 490 T; 0 U; 0 Other;
 Query Match 7.1%; Score 224.4; DB 13; Length 2117;
 Best Local Similarity 55.3%; Pred. No. 9.2e-48;
 Matches 457; Conservative 0; Mismatches 366; Indels 3; Gaps 1;
 QY 449 TTTAGAGCAAACTGGCTTCTGCTGATGCCATTCAGACTTCGTGGTTTCCAGAAC 508
 DB 714 TTCAGGAACAATGGAAGACTGCTGCGACCTGCCATCTCTTTGCTATGACTAAATG 773
 QY 509 AACACTCCAGTGGGACTTAACATGAGCTACGAGTGGAAAGCAAGAACACATCCCCATT 568
 DB 774 AATTCCCTATGGGAGAGCATGTGATGACGGGAGTTTTTATCTCATTCACCAT 833
 QY 569 CGAGAGAACATTTTCCACATGTTTCCAGTGTGCGAGCCTTTTGGACTATCCCTATAAC 628
 DB 834 GACAAATTCAACTTACTCTCTCTTCCACAGGCAACCCCAT---TCCAGCTGCCATTGAAG 890

```
QY 629 CAGTGTGAGTGTGGTAAATGGGGAATTCACAAAGTCTCTCGCGAGCAGAAATT 688
DB 891 AAATGCGCGTGTGGGAAATGGTGGATTCGAAGAGAGTGGCTGTGGCGTCAATA 950
QY 689 GATAAATCTGACTTCGTCTTCAGGTGTAACTCCCGCCCAATCACAGGAGCGCTAGTAA 748
DB 951 GATGAAGCAAAATTTGTGTCATGCATGCAATCTCCCTCTTTTCAAGTGAATACACTAAG 1010
QY 749 GATGTTGGAGCAAAACAAATCTTGAGTGTCAATCCAGCATTTATACCTGTAAGTAC 808
DB 1011 GATGTTGGATCCAAAAGTCAGTTAGTGCAGAGTAATCCAGCATAAATTCGGCAAAAGTTT 1070
QY 809 CAGAAATTTGAGCAGAGAAAGACACAGTTTGTGGAGGACATCTCCACCTATGAGATGCA 868
DB 1071 CAGAACTTCTGTGTCGAGAAAGACATTTGTGGACATGAAATCTATTAACACAGT 1130
QY 869 TTCCTCTCTCCAGCAATTTTCTATCGGCGCAACACAGGCAATCTCTTTTAAAGTCTAC 928
DB 1131 TACATCTACATGCTGCTTTTCTATGAGACAGAGACAGAGCCATCTTTGAGGTTTAT 1190
QY 929 CAAACACTCAAGAGTCAAAATGAGGCAAAAGTTCTCTTTCATCCAGGTACCTG 988
DB 1191 TATACACTGTGAGATGTTGGTCCCAATCAACAGTGTCTTTGCGCAACCCCACTTCTG 1250
QY 989 AGACACTCGCTCTTTCTGGAGAACTTAAGGGTGAAGTGCATACGCTGTGTCACAGGC 1048
DB 1251 CGTAGCATTGGAAAGTCTTGGAAATAGTAGGAATCCATGCCAAGCGCTGTGCCACAGGA 1310
QY 1049 TTGATGATTGCAAGTGTGCTGTGAACTGTGTGAAAGCTGAAGTCTACCGATTTCTGG 1108
DB 1311 CTTTTTCTGGTGCAGCGAGTCTGGGTCTGTGAGAGGTGGCCATCTATGGCTTCTGG 1370
QY 1109 CTTTTCTCTAAGCATATCGAAGACACCCCACTCAGTCAACCACTATATGATAACATGTTA 1168
DB 1371 CTTTTCTCTGTAATATGATGAGCAGCGCCATCAGCCACCACTACTATGACAACTCTTA 1430
QY 1169 CTTAGCATGTTTCCACAGATGCTTAAGATACAGCCAAATGCTCCAGTCCATATG 1228
DB 1431 CTTTTTCTGGCTTCCATGCCATGCGCGAGGAATTTCTCCAACTCTGGTATCTTCATATA 1490
QY 1229 AGAGAAATCTCAAACTGCAATTCAGCAATTCGAAAGCTGAAACGGCTTAAC 1274
DB 1491 ATCGGTGCACTCAGATGCAAGTGCAGCTGACCCATGTGAAGATACCTCAC 1536
```

RESULT 11
AAT03694

ID AAT03694 standard; cDNA to mRNA; 1704 BP.

AC AAT03694;

DT 07-AUG-1996 (first entry)

XX Human alpha 2,8-sialyltransferase cDNA.

XX Human melanoma; SK-Mel-28; ganglioside 3-synthase; tumour antigen;
KW detection; Northern blotting; glycosylation; ds.

XX Homo sapiens.

FH Key Location/Qualifiers
CDS 149..1174FT FT /*tag= a
FT FT /product= "alpha-2,8-sialyltransferase"

FT FT 721..766

FT FT /*tag= b

FT FT /note= "the sequence of bases 721-766 is missing from the
FT FT specification. Generic codons coding for the amino acid
FT FT sequence VDNMKIY- NMSIYMP have been inserted by the
FT FT indexer"

FN JP07327678-A.

XX

```
PD 19-DEC-1995.
XX 03-JUN-1994; 94JP-00145450.
XX 03-JUN-1994; 94JP-00145450.
XX (SEBK ) SEIKAGAKU KOGYO CO LTD.
PA WPI; 1996-072336/08.
DR P-PSDB; AAR88326.
XX Alpha 2,8-sialyltransferase cDNA - useful for sensitive detection of  
PT tumour antigen by Northern blotting.  
XX Claim 1; Page 5-8; 10pp; Japanese.  
XX A cDNA clone coding for alpha 2,8-sialyltransferase was isolated from a  
CC human melanoma SK-Mel-28 cDNA library and sequenced. The insert codes for  
CC a protein having ganglioside 3-synthase activity which is useful as a  
CC biochemical reagent. The cDNA is useful for detecting tumour antigen by  
CC Northern blot analysis  
XX Sequence 1704 BP; 452 A; 377 C; 434 G; 426 T; 0 U; 15 Other;
```

Query Match 7.0%; Score 222.2; DB 2; Length 1704;

Best Local Similarity 54.4%; Frsh. No. 3.1e-47;

Matches 449; Conservative 8; Mismatches 366; Indels 3; Gaps 1;

```
QY 449 TTTAGAGCAAACTGGCTTCCTCTGCGATGCGCAATTCAGAGCTTCGTGGTTTCCAGAAC 508
DB 335 TTCAGGAAACAATGGAAGACTCTGCGACCTGCCATCTCTTGTCTATGACTAAATG 394
QY 509 AACACTCCAGTGGGAGTAACTAGCTACGAGTGGAAAGCAAGAAACACATCCCATTT 568
DB 395 AATTCCTTATGGGAGAGCAATGTGTATGACGGGGAGTTTATATCTATCTATTCACCA 454
QY 569 CGAGAGCAATTTCCACATCTTCCAGTGTGCGAGCCCTTTGTGAGCTATCCCTATAC 628
DB 455 GACAATTCACATCTCTCTCTCCACAGGCAACCCCAT---TCAGCTGCATTTGAAG 511
QY 629 CAGTGTGAGTGGTGTGTAATGAGGGAATTCACAAAGTCTCTCTGCGGAGCAGAAATT 688
DB 512 AAATGCGCGTGTGGGAAATGATGGGATTCGAAGAGAGTGGCTGTGGCGTCAATA 571
QY 689 GATAAATCTGACTTCGTCTTCAATGTAACTCCCGCCCAATCACAGGAGCGCTAGTAA 748
DB 572 GATGAAGCAAAATTTGTTCATGCAATCTCCCTCTCTTGTCAAGTGAATACACTAAG 631
QY 749 GATGTTGGAGCAAAACAAATCTGTGACTCTCAATCCAGCATTTATACCTGAGTAC 808
DB 632 GATGTTGGATCCAAAGTCAAGTGTGAGCAGCTAATCCAGCATTAATTCGGCAAAAGTTT 691
QY 809 CAGAAATTTGAAGGAGAAAGAAAGACAGTTTTTGGAGGACATCTCCACCTATGAGATGCA 868
DB 692 CAGAACCTCTGTGTGTCAGAAACACATTTGNGAYATGAARATHATAAYCAYWSN 751
QY 869 TTCCTCTCTGCGCAGCATTTTCTATGGGCCCAACACAGGCAATCTCTTTTAAAGTCTAC 928
DB 752 TAYATHAYATGCTGCTTCTTTTATGAAGACAGAAACAGAGCCATCTTTGAGGGTTAT 811
QY 929 CAACACTCAAGAGATCAAAAATGAGGCAAAAGTTCTCTTCTTCCATCCAGGTACCTG 988
DB 812 TATACACTGTGAGATGTTGGTGCAATCAAAACAGTGTGTTGGCAACCCCACTTTCTG 871
QY 989 AGACACTCGCTCTTTTCTGGAACTAAAGGGTGAAGTGCATACCGCTGTGTCCACAGCG 1048
DB 872 CGTAGCATTGGAAAGTCTGGAAGTAGAGAAATCCATGCCAAGCGCTGTGCCACAGGA 931
QY 1049 TTGATGATTGCAAGTGTGCTGTGGAACTGTGTGAAACAGTGAAGCTCTACGGATTTCTG 1108
DB 932 CTTTTTCTGGTGCAGCGAGCTCTGGTCTCTGTGAAGAGGTGGCCATCTATGGCTTCTGG 991
QY 1109 CTTTTCTTAAGACTATCGAAGAACCCCACTCAGTCACCACTACTATGATAACATGTTA 1168
```

Db 992 CCCCCCTCTCTGTAATATGATGAGCAGCCCATGAGCCCACTACTATGACAACTCTTA 1051
 QY 1169 CTAAGACATGGTTTCCACAGATGCTTAAAGATACAGCCAAATGCTCCAGTCCATATG 1228
 Db 1052 CCCCCCTCTCTGTAATATGATGAGCAGCCCATGAGCCCACTACTATGACAACTCTTA 1111
 QY 1229 AGAGAACTCTCAAACTGCAATTCAGCAAAATGTGAAACGGCTTAAC 1274
 Db 1112 ATCGGTGCACTGAGAATGCACTGAGCCCACTACTATGAGATACCTCAC 1157

RESULT 12

ADBS8122
 ID ADBS8122 standard; DNA; 1223 BP.

AC ADBS8122;
 XX

DT 04-DEC-2003 (first entry)
 DE Toxicity-related gene, SEQ ID 3148.

XX Toxic; toxin; gene expression profile; hepatotoxicity; liver;
 KW drug screening; toxicity assay; ds.

XX Unidentified.
 OS

XX WO2003064624-A2.
 PN

XX 07-AUG-2003.
 PD

XX 31-JAN-2003; 2003WO-US003194.
 PF

XX 31-JAN-2002; 2002US-00060087.
 PR

XX 15-MAR-2002; 2002US-0364045P.
 PR

XX 15-MAR-2002; 2002US-0364055P.
 PR

XX 30-DEC-2002; 2002US-0436643P.
 PR

XX (GENE-) GENE LOGIC INC.
 PA

XX Mendrick D, Porter M, Johnson K, Higgs B, Castle A, Elashoff M;
 PI

XX WPI; 2003-689530/65.
 DR

XX Predicting a toxic effect of a compound, useful in identifying toxicity
 PT markers in liver tissues or cells for drug screening and toxicity assays,
 PT comprises preparing gene expression profile of tissue or cells exposed to
 PT the compound.

XX Claim 1; SEQ ID NO 3148; 1156pp; English.
 PS

XX The present invention relates to a method for predicting a toxic effect
 CC of a compound. The method comprises preparing a gene expression profile
 CC of a tissue or cell sample exposed to the compound, and comparing the
 CC gene expression profile to a database comprising SEQ ID 1-4925, where
 CC differential expression of the gene indicates at least one toxic effect.
 CC The method is useful for predicting at least one toxic effect of a
 CC compound, predicting hepatotoxicity or the progression of a toxic effect
 CC of a compound, identifying an agent that modulates the onset or
 CC progression of a toxic response, predicting the cellular pathways that a
 CC compound modulates in a cell, and identifying an agent that modulates at
 CC least one activity of a protein. The method and compositions of the
 CC present invention using a database of genes having liver toxin-induced
 CC differential expression, are useful in identifying toxicity markers in
 CC liver tissues or cells for drug screening and toxicity assays. Note: The
 CC sequence data for this patent did not form part of the printed from WIPO
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 1223 BP; 281 A; 322 C; 331 G; 289 T; 0 U; 0 Other;
 SQ

Query Match 6.5%; Score 207; DB 10; Length 1223;
 Best Local Similarity 54.2%; Pred. No. 2.6e-43;

Matches 442; Conservative 0; Mismatches 370; Indels 3; Gaps 1;
 QY 449 TTTAGAGCAAACTGGCTTCTCTTGGATGCCAATTCAGACTTCTGGTGTTCACAGAAC 508
 Db 280 TTCAGGAGACAAATGGAAGACTCTTGAACCCGCCCATCTCTTTGCAATGACGAAGGTG 339
 QY 509 AACACTCCAGTGGGAGCTAACATAGCTTACGAGTGGGAAGCAAGAAACACATCCCATTT 568
 Db 340 AACCCCCCATGGGGAAGAGCTTGGTATGACGGGAGTTTCTATCTCGCTCACATC 399
 QY 569 CGAGAGAACTTTTCCACATGTTCCAGTGTGCGAGCCCTTTTGTGGACTATCCCTTATAAC 628
 Db 400 GACAACTCCACGTACTCCCTCTTCCCCAGGCAACCCCTCT--TCCAGCTGCCATTGAAG 456
 QY 629 CAGTGTGCAAGTGGTGTGTAATGCGGGAAATTTCTCAAAAGTCTCTCTGCGGAGCAGAAATTT 688
 Db 457 AAATGTGCGGTGGTGGGAAACGCTGGGATTTCTGAAGATGAGTGGCTGTGGCGGTCAAATA 516
 QY 689 GATTAATCTGACTTCTGCTTCTCAGTGTAACTCCCTCCCAATCACAGGAGCGCTAGTAAA 748
 Db 517 GATGAAGCAAAATTTTGTGATGCTGCTTAACTCTCCCTCTTGTCAAGTGAATACACAGA 576
 QY 749 GATGTTGGAAGCAAAACAAATCTTGTGACTGTCAATCCAGCATTAATACCTTGAAGTAC 808
 Db 577 GATGTGGGTTCCAAGACTCAGTTTGTGACAGCTAACCCAGCATAAATTCGCCAGAGATTT 636
 QY 809 CAGAAATTGAAGGAGAAAGAGCAAGTTTTTGGAGGACATCTCCACCTATGGAGATGCA 868
 Db 637 GAAACCTGTGTGTTCCAGAAAGAGTTTGTGGACACATGAAGATCTATAACACAGT 696
 QY 869 TTCCTCTCTCGTCCAGCATTTTCTATCGGGCCCAACACAGGATCTCTCTTTTAAAGTCTAC 928
 Db 697 TACATATACATGCTGCTGCTTTTCTATGAAGACGGGAACAGAGCGCTCTCTCGTGTGTAC 756
 QY 929 CAACACTCTCAAGAGTCAAAATTTAGGCAAAAGGTTCTCTTCTTCCATCCCAAGTACCTG 988
 Db 757 TACACCTTGAAGACGCTGGAGCAATCAACAGTGTCTTTTGTCTAACCCCACTTTCTT 816
 QY 989 AGACACTCTGCTCTTTTCTGGAACAATAAGGGGTGACTGCATACCGTGTGTCCACAGGC 1048
 Db 817 CGAAACATTTGAAAAGTTCTGGAAGGTTAGGGGGATCCATGCCAAGCGCTGTCTACAGGA 876
 QY 1049 TTGATGATGCAAGTGTGCTGTGGAACTGTGTAACAGTGAAGCTCTACGGATTCTGG 1108
 Db 877 CTCCTTTTGGTGTGACGCTTGGGCTTTTGTGAGGAGGTGCCATCTACGGCTTCTGG 936
 QY 1109 CCTTTCTCTAAGACTATCGAAGTACCCCACTCAGTCAACCACTACTATGATAACATGTTA 1168
 Db 937 CCTTCTCCGTGAACATGCAAGTGGAACTTATCAGCCACCATTAATGATGCAACGCTGTG 996
 QY 1169 CCTAAGCATGTTTCCACCAGATGCTTAAAGAAATACGCCAAATGCTCCAGTCCATATG 1228
 Db 997 CCTTCTCAGGCTTCCATGCCATGCTTGAAGGAGTTTCTTCAACTCTGTTATCTTCATAAA 1056
 QY 1229 AGAGAACTCTCAAACTGCAATTCAGCAAAATGTGA 1263
 Db 1057 ATGGGTGCACGTGAGGATGCAAACTGCAACCAATGTA 1091

RESULT 13

AAH34115

ID AAH34115 standard; cDNA; 1511 BP

XX AAH34115;

AC AAH34115;

XX 03-SEP-2001 (first entry)

XX Human colon cancer antigen encoding cDNA SEQ ID NO:1197.

DE Human; colon cancer; colon cancer antigen; diagnosis; detection;

XX colorectal carcinoma; chromosome 12; ss.

XX Homo sapiens.

OS

GenCore version 5.1.8		Sequence 2112, A
Copyright (c) 1993 - 2006 Bioceleration Ltd.		Sequence 6236, Ap
		Sequence 1959, Ap
		Sequence 1, Appli
		Sequence 181, App
		Sequence 183, App
		Sequence 185, App
		Sequence 187, App
		Sequence 189, App
		Sequence 191, App
		Sequence 193, App
		Sequence 195, App
		Sequence 197, App
		Sequence 199, App
		Sequence 201, App
		Sequence 1829, Ap
		Sequence 29326, A
		Sequence 748151, A
		Sequence 748152, A
		Sequence 748153, A
		Sequence 748152, A
		Sequence 1105, Ap
		Sequence 13566, A
		Sequence 34255, A
		Sequence 34256, A
		Sequence 34255, A
		Sequence 34256, A
		Sequence 2204, Ap
		Sequence 2112, A
		Sequence 6236, Ap
		Sequence 1959, Ap
		Sequence 1, Appli
		Sequence 181, App
		Sequence 183, App
		Sequence 185, App
		Sequence 187, App
		Sequence 189, App
		Sequence 191, App
		Sequence 193, App
		Sequence 195, App
		Sequence 197, App
		Sequence 199, App
		Sequence 201, App
		Sequence 1829, Ap
		Sequence 29326, A
		Sequence 748151, A
		Sequence 748152, A
		Sequence 748153, A
		Sequence 748152, A
		Sequence 1105, Ap
		Sequence 13566, A
		Sequence 34255, A
		Sequence 34256, A
		Sequence 34255, A
		Sequence 34256, A
		Sequence 2204, Ap
		Sequence 2112, A
		Sequence 6236, Ap
		Sequence 1959, Ap
		Sequence 1, Appli
		Sequence 181, App
		Sequence 183, App
		Sequence 185, App
		Sequence 187, App
		Sequence 189, App
		Sequence 191, App
		Sequence 193, App
		Sequence 195, App
		Sequence 197, App
		Sequence 199, App
		Sequence 201, App
		Sequence 1829, Ap
		Sequence 29326, A
		Sequence 748151, A
		Sequence 748152, A
		Sequence 748153, A
		Sequence 748152, A
		Sequence 1105, Ap
		Sequence 13566, A
		Sequence 34255, A
		Sequence 34256, A
		Sequence 34255, A
		Sequence 34256, A
		Sequence 2204, Ap
		Sequence 2112, A
		Sequence 6236, Ap
		Sequence 1959, Ap
		Sequence 1, Appli
		Sequence 181, App
		Sequence 183, App
		Sequence 185, App
		Sequence 187, App
		Sequence 189, App
		Sequence 191, App
		Sequence 193, App
		Sequence 195, App
		Sequence 197, App
		Sequence 199, App
		Sequence 201, App
		Sequence 1829, Ap
		Sequence 29326, A
		Sequence 748151, A
		Sequence 748152, A
		Sequence 748153, A
		Sequence 748152, A
		Sequence 1105, Ap
		Sequence 13566, A
		Sequence 34255, A
		Sequence 34256, A
		Sequence 34255, A
		Sequence 34256, A
		Sequence 2204, Ap
		Sequence 2112, A
		Sequence 6236, Ap
		Sequence 1959, Ap
		Sequence 1, Appli
		Sequence 181, App
		Sequence 183, App
		Sequence 185, App
		Sequence 187, App
		Sequence 189, App
		Sequence 191, App
		Sequence 193, App
		Sequence 195, App
		Sequence 197, App
		Sequence 199, App
		Sequence 201, App
		Sequence 1829, Ap
		Sequence 29326, A
		Sequence 748151, A
		Sequence 748152, A
		Sequence 748153, A
		Sequence 748152, A
		Sequence 1105, Ap
		Sequence 13566, A
		Sequence 34255, A
		Sequence 34256, A
		Sequence 34255, A
		Sequence 34256, A
		Sequence 2204, Ap

ALIGNMENTS

```

RESULT 1
US-10-501-930-2
; Sequence 2, Application US/10501930
; Publication No. US20060057696A1
; GENERAL INFORMATION:
; APPLICANT: TAKASHIMA, Shou
; APPLICANT: TSUJIMOTO, Masaafumi
; APPLICANT: TSUJII, Shuichi
; TITLE OF INVENTION: GLYCOSYLATING ENZYME
; FILE REFERENCE: P25687
; CURRENT APPLICATION NUMBER: US/10/501,930
; CURRENT FILING DATE: 2004-07-29
; PRIOR APPLICATION NUMBER: PCT/JP03/00883
; PRIOR FILING DATE: 2003-01-30
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 2
; LENGTH: 3166
; TYPE: DNA
; ORGANISM: Mouse
US-10-501-930-2

```

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3166	100.0	3166	12	US-10-501-930-2
2	930.2	29.4	1500	12	US-10-501-930-4
3	278.2	8.8	810	12	US-10-301-480-562554
c 3	278.2	8.8	810	12	US-10-301-480-562554
c 4	278.2	8.8	810	12	US-10-301-480-1175963
5	254	8.0	672	6	US-10-027-632-44555
6	254	8.0	672	7	US-10-027-632-44555
7	224.4	7.1	1071	7	US-10-085-117-66
8	224.4	7.1	1071	8	US-10-384-339C-136
9	224.4	7.1	2117	7	US-10-172-118-859
10	224.4	7.1	2117	7	US-10-085-117-65
11	224.4	7.1	2117	7	US-10-430-325-1
12	224.4	7.1	2117	8	US-10-342-887-859
13	210	6.6	1026	7	US-10-085-117-63
14	210	6.6	1388	7	US-10-085-117-62
15	206.6	6.5	1223	16	US-11-136-527-2140
16	175.4	5.5	1511	6	US-10-106-698-1207
17	131.4	4.2	448	3	US-09-864-761-15072

Db	181	GTGATGGAGGAGCAGAGAGGACACCAAGTGGTACCTCAGCTGCATCGACCTGAAGACACTCTG	240
Qy	241	GAGCCCGCAACCCCGGTGACACGCAACGAGAAAGCAATATCTGGATGAGAGACAAC	300
Db	241	GAGCCCGCAACCCCGGTGACACGCAACGAGAAAGCAATATCTGGATGAGAGACAAC	300
Qy	301	CCAAATAACAGAGAAATGCAAGAGTCTGCAATATAGCTTGAACCTCTTTATCTAACAAAC	360
Db	301	CCAAATAACAGAGAAATGCAAGAGTCTGCAATATAGCTTGAACCTCTTTATCTAACAAAC	360
Qy	361	GAGCGGTACTCTGAGGATGACTACCTCCAGACCAATCAAAACATACAGAGATGCCCATG	420
Db	361	GAGCGGTACTCTGAGGATGACTACCTCCAGACCAATCAAAACATACAGAGATGCCCATG	420
Qy	421	GAAACCGGACAGAGAAATATGCAATTTTAGAGCAAAACCTGGCTTCTGTTGGATGTC	480
Db	421	GAAACCGGACAGAGAAATATGCAATTTTAGAGCAAAACCTGGCTTCTGTTGGATGTC	480
Qy	481	CATTCAAGACTTCGTGGTTTCCAGAAACAACCTCCAGTGGGACTAAACATGAGCTACGA	540
Db	481	CATTCAAGACTTCGTGGTTTCCAGAAACAACCTCCAGTGGGACTAAACATGAGCTACGA	540
Qy	541	GGTGGAAAGCAAGAAACACATCCCATTCGAGAGAAACATTTTCCAGTGTGTC	600
Db	541	GGTGGAAAGCAAGAAACACATCCCATTCGAGAGAAACATTTTCCAGTGTGTC	600
Qy	601	GCAGCTTTTGTGACTATCCCTATAACACAGTGTGCACTGGTGTGTAATGGGGAAATCT	660
Db	601	GCAGCTTTTGTGACTATCCCTATAACACAGTGTGCACTGGTGTGTAATGGGGAAATCT	660
Qy	661	CAACAAAGTCTCTCGCGAGCAGAAATGATAATCTGACTTCGTCTCAGTGTGTAACCT	720
Db	661	CAACAAAGTCTCTCGCGAGCAGAAATGATAATCTGACTTCGTCTCAGTGTGTAACCT	720
Qy	721	CCCCCAATCACAGGAGCGCTAGTAAAGATGTTGGAAAGCAAAACAAATCTTTGACTGT	780
Db	721	CCCCCAATCACAGGAGCGCTAGTAAAGATGTTGGAAAGCAAAACAAATCTTTGACTGT	780
Qy	781	CAATCCAGCAATTAACCCCTGAAGTACAGAAATTTGAAGAGAGAAAGACAGTGTCTT	840
Db	781	CAATCCAGCAATTAACCCCTGAAGTACAGAAATTTGAAGAGAGAAAGACAGTGTCTT	840
Qy	841	GGAGACATCTCCACTATGAGATGCATCTCCTCCTCGCCAGCATTTTCTATCGGCG	900
Db	841	GGAGACATCTCCACTATGAGATGCATCTCCTCCTCGCCAGCATTTTCTATCGGCG	900
Qy	901	CAACACAGGCATCTCTTTTAAAGTCTACCAACACTCAAGAGTCAAAATGAGGCAAAA	960
Db	901	CAACACAGGCATCTCTTTTAAAGTCTACCAACACTCAAGAGTCAAAATGAGGCAAAA	960
Qy	961	GGTCTCTTCTTCATCCAGGTACTGAGACACCTCGCTCTTTCTGGAGAACTAAAGG	1020
Db	961	GGTCTCTTCTTCATCCAGGTACTGAGACACCTCGCTCTTTCTGGAGAACTAAAGG	1020
Qy	1021	GGTGACTGCATACCGTGTCCACAGGCTTGATGATTCAGTGTGCTGTGGAACTGTG	1080
Db	1021	GGTGACTGCATACCGTGTCCACAGGCTTGATGATTCAGTGTGCTGTGGAACTGTG	1080
Qy	1081	TGAACACGTGAAGCTCTACGGATTTCTGCGCTTTCTCTAAGACTATCGAAGACACCCCACT	1140
Db	1081	TGAACACGTGAAGCTCTACGGATTTCTGCGCTTTCTCTAAGACTATCGAAGACACCCCACT	1140
Qy	1141	CAGTCAACACTACTATGATAACATGTTACCTAAGCATGGTTTCCACAGATGCCTTAAGA	1200
Db	1141	CAGTCAACACTACTATGATAACATGTTACCTAAGCATGGTTTCCACAGATGCCTTAAGA	1200
Qy	1201	ATACAGCCAAATGCTCCAGCTCCATATCAGAGGATCCTCAAACTGCAATTCAGCAATG	1260
Db	1201	ATACAGCCAAATGCTCCAGCTCCATATCAGAGGATCCTCAAACTGCAATTCAGCAATG	1260
Qy	1261	TGAACGCGCTTAAAGCTTTCTAGAGAGGAAATTAATTTTCAGGAGGTGGAGTGTGTCA	1320
Db	1261	TGAACGCGCTTAAAGCTTTCTAGAGAGGAAATTAATTTTCAGGAGGTGGAGTGTGTCA	1320
Qy	1321	CAGCATCTCCAAAAGCCAAATGAAAGGACACAGAGAAAGCATGAATTAACAAGGCGCT	1380
Db	1321	CAGCATCTCCAAAAGCCAAATGAAAGGACACAGAGAAAGCATGAATTAACAAGGCGCT	1380
Qy	1381	CTCCCACTTGTCTAGACCAAAAGCACCGGCCCCCACTCACCTTTGAGCTCTCACAGTGCA	1440
Db	1381	CTCCCACTTGTCTAGACCAAAAGCACCGGCCCCCACTCACCTTTGAGCTCTCACAGTGCA	1440
Qy	1441	CTCAATCTCACTTCAACGGTCTCTCTGAGAATAGAGACCAAAACATACAGCTTGGAT	1500
Db	1441	CTCAATCTCACTTCAACGGTCTCTCTGAGAATAGAGACCAAAACATACAGCTTGGAT	1500
Qy	1501	AAGTAAATAGAGATAATTTTTCATCATATAGAAATTTGATTTGAGCAGGGTCTCTCA	1560
Db	1501	AAGTAAATAGAGATAATTTTTCATCATATAGAAATTTGATTTGAGCAGGGTCTCTCA	1560
Qy	1561	GAATGCTTCTTGTTCCTATCCATGATAGCCATTTCCACCTTTATCAGAGTGGTAATGAA	1620
Db	1561	GAATGCTTCTTGTTCCTATCCATGATAGCCATTTCCACCTTTATCAGAGTGGTAATGAA	1620
Qy	1621	ACTGTCAATTTGTGCCAAAGACCTTTCTGAAGAGAAATGTCTGAATCATGCGCGAGTTT	1680
Db	1621	ACTGTCAATTTGTGCCAAAGACCTTTCTGAAGAGAAATGTCTGAATCATGCGCGAGTTT	1680
Qy	1681	TTACACACAGCTCTTCTTTTATTAATAATCTCTCCATTCCTCCTCTAGTAGAGTACA	1740
Db	1681	TTACACACAGCTCTTCTTTTATTAATAATCTCTCCATTCCTCCTCTAGTAGAGTACA	1740
Qy	1741	GAACAAAAATACCTTCATGATTAAGAAAGAAAGTCTTTTACTTTAGCAATGTGCGCTG	1800
Db	1741	GAACAAAAATACCTTCATGATTAAGAAAGAAAGTCTTTTACTTTAGCAATGTGCGCTG	1800
Qy	1801	CTTCTGATTCAGTTTCGCTTGTGTAATTAAGCTGGGTTGGGTTTGGTGGATTTGGGCG	1860
Db	1801	CTTCTGATTCAGTTTCGCTTGTGTAATTAAGCTGGGTTGGGTTTGGTGGATTTGGGCG	1860
Qy	1861	GTTCCTTCACTCTTTTGTCTAATTTTCTTACCTTTATCAGTTTGTATTCGAGTCTCC	1920
Db	1861	GTTCCTTCACTCTTTTGTCTAATTTTCTTACCTTTATCAGTTTGTATTCGAGTCTCC	1920
Qy	1921	TGCTTTGGGATTTGCAATTTCTCCTCCACTGACAGGATCAACTCAATGACATAAAGTA	1980
Db	1921	TGCTTTGGGATTTGCAATTTCTCCTCCACTGACAGGATCAACTCAATGACATAAAGTA	1980
Qy	1981	GTTCAAAACATCCATTTGCTCTCATGTTTATCCATAAAGTTACTCATCTGATTTTAT	2040
Db	1981	GTTCAAAACATCCATTTGCTCTCATGTTTATCCATAAAGTTACTCATCTGATTTTAT	2040
Qy	2041	TAAATAGTGAACATCTACTTGTATCAGACCCGAGGACCATCTCCATTCGAGAAATG	2100
Db	2041	TAAATAGTGAACATCTACTTGTATCAGACCCGAGGACCATCTCCATTCGAGAAATG	2100
Qy	2101	AAGATATTGTCACCTGGCAGAAATTCAGGTGTGTCCTAATTAATGATAAGATACCAAGC	2160
Db	2101	AAGATATTGTCACCTGGCAGAAATTCAGGTGTGTCCTAATTAATGATAAGATACCAAGC	2160
Qy	2161	ATCATCATGCGAGTTATGAACAAGTCTGAAAGGATCATAGACAGGGGTGTTAAATCT	2220
Db	2161	ATCATCATGCGAGTTATGAACAAGTCTGAAAGGATCATAGACAGGGGTGTTAAATCT	2220
Qy	2221	GATCCAGTAAATTAACCTTCACTGACCTAATTTTTCAGGAAAGATTAATTTTCAAAATAA	2280
Db	2221	GATCCAGTAAATTAACCTTCACTGACCTAATTTTTCAGGAAAGATTAATTTTCAAAATAA	2280
Qy	2281	AACTAGTAAATGAACCAATTTTTCAGGCAATTAAGTGGATTTCTGAGTAAAGAAAGGAA	2340
Db	2281	AACTAGTAAATGAACCAATTTTTCAGGCAATTAAGTGGATTTCTGAGTAAAGAAAGGAA	2340
Qy	2341	CAGCAGGAGAAAGCTGTTCGCTTGGTCTGATTAACCAATATGAGCATGCTGGAAGGAGT	2400
Db	2341	CAGCAGGAGAAAGCTGTTCGCTTGGTCTGATTAACCAATATGAGCATGCTGGAAGGAGT	2400


```

Db      810 GTACCTGAAGATGCGCCCTTTCTGGAGAACTAAGGTGTGACTGCATACGGCTTGTG 751
QY      1042 CACAGGCTTGATGATGCAAGTGTGCGTGTGAACTGTGTGAAACGTGAAGCTTACGG 1101
Db      750 CACCGGCTTGATGATGATGCAAGTGTGCGTGTGAACTGTGTGAAACGTGAAGCTTACGG 691
QY      1102 ATTCCTGGCTTCTCTAGACTATCGAAGACACCCCACTCAGTCACCACTACTATGATA 1161
Db      690 ATTCCTGGCTTCTCTAGAACTGTGAAGACATACCTGTGACGCACTCATATATGACAA 631
QY      1162 CATGTTACCTAAGCATGTTTCCACAGATGCTTAAAGAAATACAGCAAAATGCTCCAGCT 1221
Db      630 CAACTACTAAACATGTTTCCATCAGATGCCAAAGATACAGCCAGATCTTCAACT 571
QY      1222 CCATATGAGAGAAATCTCTAAATCTGCAATTCAGCAAAATGTGAACGGCTT-----AACGTT 1277
Db      570 TCACATGAAGAAATCTCTAAATCTGCAATTTAGCAAAATGTGAAGTGCCTTAAACAAAGTA 511
QY      1278 TCTTAGAGGAAATATTTTCAGGA-GGTGGAGTGGATGTGTACAGATCTTCCAAAAG 1336
Db      510 TCTTAAATGGAATATTTTAAATATATGCGATAGGTATTAACAATGTCTTCCAAACAC 451
QY      1337 CCAATAGAAGAGGACACAGAGAAAGCATGAATTAACAAGCGCTCTCCCACTTGTCTAGA 1396
Db      450 C-----AAGGAGGTGGCTAAGAGATATTTTGAAGTACGACCCCAAAATTTGGTTGA 399
QY      1397 CCAAGCCACCGCCCACTCACTTTGCGAGCTTCCAGAGTCACTC-ATTCCTCACCTTC 1455
Db      398 CCAAG-----CTTCCCACTCATTTTGCATGATGGCAAGTCAATCAATCTTCTCATC 344
QY      1456 AACGTTCTTCTGAGAAATAGAGACCAAAACATCAGACTTGGATAGTAAGTAAGTACGATA 1515
Db      343 TTCAATTTTCTCTTAAACATGAGACCAATATCTGACTCATACAAATTTTATAATTT 284
QY      1516 ATTTTTCATCATCATAGAAATTTGATTTGAGCCAGGGTCTCTCAGAAATGCTTCTGTT 1575
Db      283 ATGAAATTTTGGATGCGCTTTTGGTGGAGTGGAGTCAAAAGATATTTCTTCTT 224
QY      1576 CCTATCCATGATGACCATTTCCACCTTTATCAGAGTGGTAATGAACCTGTGCAATTTGTC 1635
Db      223 ACTATTTATACAGCCATCTCCACCTTCCACAGACCGATATAAAGCATGCTTCTTG- 165
QY      1636 CAAAGACCTTTCTGAAGAGAAATGTCTGAAT-----CATGGCGGAGTTTTCACACA 1689
Db      164 --GAGATCTTCTTAAAGAGAAATGCTGCATTTACCAGTATCCATTTGGATTTCTTCTGT 107
QY      1690 GCTCTTCTTTATAATAATCTTCCCAATTTCCCTCTCTAGTAGTACAGAAACAAA 1749
Db      106 GCAGTCTCTATCAATATATATCTTCCCATTTTCTGTCTCAAGAGCTGCATGCAAGAA 47
QY      1750 TACCCCTGATGATTCAGGAAG 1770
Db      46 ATTATATATTAAGCAAAAG 26

```

```

RESULT 5
US-10-027-632-44555
; Sequence 44555, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; PENDING FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29

```

```

; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/187,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/186,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/186,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 44555
; LENGTH: 672
; TYPE: DNA
; ORGANISM: Human
; US-10-027-632-44555

Query Match      8.0%; Score 254; DB 6; Length 672;
Best Local Similarity 67.4%; Pred. No. 1.8e-54;
Matches 437; Conservative 1; Mismatches 191; Indels 19; Gaps 5;

QY      993 ACCTCGCTCTTTTCTGGAGAACTAAAGGGTGACTGCATACCGCTGTCCACAGGCTTGA 1052
Db      33 ATCTGGCCCTTTTCTGGTAAACAAAGGTGTGACTGCATACCGCTGTCCACCGGCTTGA 92
QY      1053 TGATTGCAAGTGTGCTGTGGAACTGTGTGAAACGTGAAGCTCTACGATTTCTGGCCCTT 1112
Db      93 TGATCACAAGTGTGTGCACTGGAACTGTGTAAATGTGAAGCTGTATGGATTTCTGGCCCT 152
QY      1113 TCTTAAGACTATCGAAGACACCTCACTCAGTCACCACTACTATGATACATGTTTACCTTA 1172
Db      153 TCTTAAACTGTAGAAGACATCTGTGAGCCATCACTATTTATGACAACAAGCTACCTTA 212
QY      1173 AGCATGTTTCCACAGATGCTTAAAGAAATCAGCAAAATGCTCCAGTCTCATATGAGAG 1232
Db      213 AACATGTTTCCATCAGATGCTTAAAGAAATCAGCCAGATCTCTCAACTTCCATGAAAG 272
QY      1233 GAATCCTCAAACTGCAATTCAGAAATCTGAAACGGCTT---AAGTCTTCTTAGAAGGA 1288
Db      273 GAATCCTCAAACTGCAATTTAGAAATGTGAAGTGCCTTAAACAAGTATCTTAAATGG 332
QY      1289 GAATAAATTTCAAGGA-GGTGGAGTGTGATGTGCACAGCATCTCCAAAAGCCAAATAGAAGA 1347
Db      333 GAATAAATTTAATAATATGTCAGTGGTGAATTAACAATGCTCTCCAAACACC-----AA 384
QY      1348 AGGCACAGAGAAAGCATGAATTAAGAGGCTCTCCCACTTGTCTAGACCAAAAGCCACC 1407
Db      385 AGGAGTGGCTTAAAGAGTATTTTGAATGAGCGCCCAAAATTTGGTTTGACCAAG----- 439
QY      1408 CGCCCCCACTCACTTTGCGAGCTTCCAGAGTCACTC-ATTCTCAGCTTCAAGCTTCTTTC 1466
Db      440 CTTCCCACTCATTTTGCATGATGGCAAGTCAATCAATCTCTCTCATCTTCTTCTTTC 499
QY      1467 TCTGAGAAATAGAGACCAAAACATTCAGACTTGATAAGTAAATAGATAATTTTCAAAAT 1526
Db      500 TCTTATTAACATGAGACCACTATTTGACTCATCAAAATTTTATATTTATGAAATTAAT 559
QY      1527 CATCATAGAATTTGATTTGAGCAAGGCTCTCTCAGAAATGCTTCTTGTCTTATCATGA 1586
Db      560 TGGCATGGCCCTTTGGTGTGAGGTAGGAGCTCAAGAAATATTTCTTTCTTACTATTATAC 619
QY      1587 TAGCCATTTCCACCTTTATCATGAGTGGTAATGAAGTGTGCAATTTGTG 1634
Db      620 CAGCCATCTCCACCTTCCACAGTCCGATATTAAGAGCATGCTTCTTCTTG 667

```

```

RESULT 6
US-10-027-632-44555
; Sequence 44555, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; Polymorphisms in the Human Genome

```

FILE REFERENCE: 108827.129
CURRENT APPLICATION NUMBER: US/10/027,632
CURRENT FILING DATE: 2002-04-30
PRIORITY APPLICATION NUMBER: US 60/218,006
PRIORITY FILING DATE: 2000-07-12
PRIORITY APPLICATION NUMBER: US 60/198,676
PRIORITY FILING DATE: 2000-04-20
PRIORITY APPLICATION NUMBER: US 60/193,483
PRIORITY FILING DATE: 2000-03-29
PRIORITY APPLICATION NUMBER: US 60/185,218
PRIORITY FILING DATE: 2000-02-24
PRIORITY APPLICATION NUMBER: US 60/167,363
PRIORITY FILING DATE: 1999-11-23
PRIORITY APPLICATION NUMBER: US 60/156,358
PRIORITY FILING DATE: 1998-09-28
PRIORITY APPLICATION NUMBER: US 60/146,002
PRIORITY FILING DATE: 1999-08-09
NUMBER OF SEQ ID NOS: 325720
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 44555
LENGTH: 672
TYPE: DNA
ORGANISM: Human
US-10-027-632-44555

Query Match 8.0%; Score 254; DB 7; Length 672;
Best Local Similarity 67.4%; Pred. No. 1.8e-54;
Matches 437; Conservative 1; Mismatches 191; Indels 19; Gaps 5;
QY 993 ACCTCGCTCTTTCTGAGAACTAAAGGGGTGACTGTCATACCGCTGTGCCAGGCTTGA 1052
DB 33 ATCTGGCCCTTCTGTTAGAACTAAAGGTGTGACTGTCATACCGCTGTGCCAGGCTTGA 92
QY 1053 TGATTGCAAGTGTCTGTGGAACCTGTGTGAAACGTGAAAGCTCTACGGATTCTGGCCCTT 1112
DB 93 TGATCACAAGTGTTCGACGTGGAACCTGTGTAAAGATGTGAAAGCTGTATGGATTCTGGCCCT 152
QY 1113 TCTCTAAGACTATCAAGACACCCCACTCAGTCACCACTACTATGATTAACATGTTTACCTA 1172
DB 153 TCTCTAAGACTATCAAGACACCACTCAGTCACCACTACTATGATTAACATGTTTACCTA 212
QY 1173 AGCATGGTTTCCACAGATGCTTAAAGAAATACAGCAAAATGCTCCAGCTCCATATGAGAG 1232
DB 213 AACATGGTTTCCATCAGATGCCAAGAAATACAGCCAGATCTTCAACTTCACATGAAG 272
QY 1233 GAATCCTCAAACTGCAATTTAGCAAAATGTGAAACGGCTT---AACGTTTCTTAGAAGGA 1288
DB 273 GAATCCTCAAACTGCAATTTAGCAAAATGTGAAAGTGGCTTAAACAAAGTATCTTAAATGG 332
QY 1289 GAATAATTTTCAGGA-GGTGGAGTGGATGTCACAGCATCTCCAAAGCCCAATAGAGA 1347
DB 333 GAATAATTTTAAATAATATGAGTAGGTGATTAACAAATGTCTCCAAACACC-----AA 384
QY 1348 AGGCACAGAGAAGCATGAATTAACAAAGCGCTCTCCCACTTGTCTAGACCAAAAGCCACC 1407
DB 385 AGGAGTGGCTTAAGAGTATTTGAGTAGGCCCAAAATTTGGTTGACCAAG-----439
QY 1408 CGCCCCCACTCACTTTGAGGCTTCCACGAGTCACTC-ATTCCTACCTTTCAACGTTCTTTTC 1466
DB 440 CTTCCCACTCATTTTGAATGATGGCAAGTCATTCATCTCTCTCATCTTCAITTTTTTC 499
QY 1467 TCTGAGATAGAGACCAAAACATCAGACTTGGATTAAGTAAATGAGATAATTTTCAAT 1526
DB 500 TCTTTAATCAATGGACACCATATCTGACTCATACAAATTTTATTAATTTATGAATAAT 559
QY 1527 CATCATGAATTTGATTTGAGCGAGGTCTCTCAGAATGCTTCTTGTCTCTATCCATGA 1586
DB 560 TGGCATGGCTTTGGTGTGAGTAGGAGTCAAGAAATATTTCTTCTTCTTACTATTTATAC 619
QY 1587 TAGCCATTTCCACCTTTATCATCAGAGTGGTAATGAACCTGTGCAATTTGTG 1634
DB 620 CAGCCATCTCCACCTTTCACAGACCGATATTTAAAGCATGCTTCTTCTTG 667

RESULT 7
US-10-085-117-66
Sequence 66, Application US/10085117
Publication No. US200302334A1
GENERAL INFORMATION:
APPLICANT: Morris, David W.
APPLICANT: Engelhard, Eric K.
TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR CANCER
FILE REFERENCE: 52945200121
CURRENT APPLICATION NUMBER: US/10/085,117
CURRENT FILING DATE: 2002-02-27
PRIORITY APPLICATION NUMBER: US 09/739,586
PRIORITY FILING DATE: 2001-03-02
NUMBER OF SEQ ID NOS: 361
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 66
LENGTH: 1071
TYPE: DNA
ORGANISM: Homo sapiens
US-10-085-117-66

Query Match 7.1%; Score 224.4; DB 7; Length 1071;
Best Local Similarity 55.3%; Pred. No. 1.2e-46;
Matches 457; Conservative 0; Mismatches 366; Indels 3; Gaps 1;
QY 449 TTTAGAGCAAACTGGCTTCTCTTGGATGCCATTCAGACTTGGTGGTTTCCAGAAC 508
DB 232 TTCAGGAAACAAATGGAAGACTCTGCGACCTCCCATCTCTTTGCTATGACTAAATG 291
QY 509 AACACTCCAGTGGGACTTAACATTAAGCTTACAGGTGGAAGCAAGAAACACATCCCCATT 568
DB 292 AATTCCCTATGGGGAAGAGCAATGTTATGACGGGAGTTTTTATATCTATTCACCAT 351
QY 569 CGAGAGAACATTTTCCACATGTTTCCAGTGTCCAGGCTTTTGTGAGATATCCCTATAAC 628
DB 352 GACAATTCAACTTACTCTCTCTCCACAGGCAACCCCAT---TCCAGCTGCCATTGAAG 408
QY 629 CAGTGTGAGTGGTGTGTAATGAGGAAATTTCTCAACAAGTCTCTTGGCGGAGCAAAAT 688
DB 409 AAATGCGCGGTGGTGGGAAATGAGGGAATTTCTGAAGAAAGTGGTGTGGCGGTCAATA 468
QY 689 GATAAATCTGACTTCGCTTCTCAGCTGTAACTCCCTCCCAATCACAGGAGCGCTAGTAAA 748
DB 469 GATGAGCAAAATTTTGTTCATGCAATGCAATCTCCCTCTTGTCAAGTGAATACACTAAG 528
QY 749 GATGTTGGAAGCAAAACAAATCTGTGACTGTCAATCCAGCATTAATAACCTGAGTAC 808
DB 529 GATGTTGGATCCAAAGTCAGTTAGTGACAGCTAATCCAGCATAAATTCGGCAAGGTTT 588
QY 809 CAGAAATTTGAAGGAGAAAGCAAGTCTTTTGGAGGACATCTCCACTATGGAGATGCA 868
DB 589 CAGAACTCTGTGGTCCAGAAACATTTGTGGAACCAATGAAATCTATAACCAAGT 648
QY 869 TTTCTCTCTCTGCGCAGCATTTCTTATCGGCGCAACACAGGCACTCTCTTTTAAAGTCTAC 928
DB 649 TACATCTACATGCTGCTCTTTCTATCAAGACAGGACAGAGCCATCTTTGAGGTTTAT 708
QY 929 CAACACTCAAGAGTCAAAAATGAGGCAAAAGGTTCTCTTTTCCATPCCAGGTACTG 988
DB 709 TATACATGTCAGATGTTGGTGTCAATCAAAACAGTGTGTTTGGCAACCCCAACTTTCTG 768
QY 989 AGACACTCTCTCTTTCTGGAACAACTAAAGGGGTGACTGCATACCGCTTGTCCACAGGC 1048
DB 769 CGTAGCAATTTGGAAGATTTCTGGAAAGTAGAGGAATCCATGCAAGCGCTGTCCACAGGA 828
QY 1049 TTGATGATTCAGAGTGTGGCTGTGAACTGTGTAAGAAAGTGTCTACGGAATTTCTGG 1108
DB 829 CTTTTTCTGGTAGCGCAGCTCTGGGTCTCTGTGAAGAGGTGGCCATCTATGGCTTCTG 888
QY 1109 CTTTTTCTTAAGACTATCGAAGTACCCCACTCAGTCAACCACTACTATGATAACATGTTA 1168
DB 889 CTTTTTCTGTGAATATGCAATGAGCAGCCCACTCAGCCCACTACTATGACAAAGCTCTTA 948

QY 1169 CCTAAGCATGTTTCCACAGATGCTTAAGAAATACAGCCAAATGCTCCAGCTCATATG 1228
Db 949 CCTTTTCTGGCTTCCATGCGCATGCCAGGAAATTTCTCCAACTCTGGTATCTTTCATAAA 1008
QY 1229 AGAGGAATCTCAAACTGCAATTCAGCAAAATGTGAACGGCTTAAC 1274
Db 1009 ATCGGTGCACTGAGNATGCACTGGACCCATGTGAAGATACCTCAC_1054

RESULT 8

US-10-384-339C-136
; Sequence 136, Application US/10384339C
; Publication No. US20040175703A1
; GENERAL INFORMATION:
; APPLICANT: Kreutzer, Roland
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR INHIBITING EXPRESSION OF A TARGET GE
; FILE REFERENCE: 20200/2002
; CURRENT APPLICATION NUMBER: US/10/384,339C
; PRIOR FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: PCT/EP02/00152
; PRIOR FILING DATE: 2002-01-09
; PRIOR APPLICATION NUMBER: DE 10100586.5
; PRIOR FILING DATE: 2001-01-09
; PRIOR APPLICATION NUMBER: DE 10155280.7
; PRIOR FILING DATE: 2001-10-26
; PRIOR APPLICATION NUMBER: DE 10158411.3
; PRIOR FILING DATE: 2001-11-29
; PRIOR APPLICATION NUMBER: DE 10160151.4
; PRIOR FILING DATE: 2001-12-07
; NUMBER OF SEQ ID NOS: 173
; SOFTWARE: PatentIn ver. 2.1
; SEQ ID NO 136
; LENGTH: 1071
; TYPE: DNA
; ORGANISM: Homo sapiens
; PUBLICATION INFORMATION:
; TITLE: GD3 synthase
; PATENT DOCUMENT NUMBER: NM003034
US-10-384-339C-136

Query Match 7.1%; Score 224.4; DB 8; Length 1071;
Best Local Similarity 55.3%; Pred. No. 1.2e-46;
Matches 457; Conservative 0; Mismatches 366; Indels 3; Gaps 1;
QY 449 TTTAGACAAACTGGCTTCTGTTGCCATGCCATTCAGACTTCGTGGTTTCCAGAAC 508
Db 232 TTCAGGAAACAATGGAGACTGTGCGACCTGCGCATCTCTTGTCTATGACTAAATG 291
QY 509 AACACTCCAGTGGGACTTAACATGAGCTACGAGTGGAAAGCAAGAAACACATCCCAT 568
Db 292 AATTCCTCATGGGAGAGCATGTGGTATGACGGGGAGTTTATATCTCATTCACCAT 351
QY 569 CGAGAGAACATTTTCCACATGTTTCCAGTGTGCGAGCTTTTGTGGACTATCCCTATAC 628
Db 352 GACAAATCAACTTACTCTCTTCTCCACAGGCAACCCCAT---TCCAGCTGCCATTGAAG 408
QY 629 CAGTGTGCAGTGGTGTGTAATGGGGAAATTCACAAAGTCTCTCTGGGAGCAGAAATT 688
Db 409 AATATCGCGGTGGTGGGAAATGGTGGGATTCGAGAGAGAGTGGTGTGGCGGTCAATA 468
QY 689 GATAAATCTGACTTCGTCTTTCAGGTGTAACTTCCCTCCCAATCACAGGAGCGCTAGTAA 748
Db 469 GATGAGCAAAATTTGTTCATGCGATGCAATCTCCCTCTTGTGTCAGTGAATACACTAAG 528
QY 749 GATGTTGAAGCAAAACAAATCTTGTGACTGTCAATCCAGCATATACCTCAAGTAC 808
Db 529 GATGTTGGATCCAAAGTCACTGTAGTGACAGCTAATCCCGACATAATTCGGCAAGGTT 588
QY 809 CAGAAATTTGAAGGAGAAAGACACAGTTTGTGGAGGACATCTCCACCTATGGAGATCA 868
Db 589 CAGAACTTCTGTGTGCCAGAAAGACATTTGTGGACAAACATGAAATCTATAACCACT 648

QY 869 TTCTCTCTCTGCGCAGCATTTTCTATCGGSCCAACACAGGCATCTCTTTTAAAGTCTAC 928
Db 649 TACATCTACATGCTGCTTTTCTATGAAGACAGGAACAGAGCCATCTTTTGGGGTTTAT 708
QY 929 CAAACACTCAAGAGAGTCAAAAATAGGCAGAAAGTTCTCTTTTCCATCCCAGGTACCTG 988
Db 709 TATACACTGTCTAGATGTTGGTGCAATCAACAGTGTCTTTTGGCAACCCCAACTTTCTG 768
QY 989 AGACACCTCTCTTTTCTGGAGCACTAAAGGGGTGACTGCATACCGCTTCTCCACAGC 1048
Db 769 CGTAGCATTTGGAAGTTCTGGAAAGTAGAGGAATCCATGCCAAGCGCTGTGCCACAGA 828
QY 1049 TTGATGATTGCAAGTGTGCTGTGCAACTGTGTGAAAACGTGAAGCTCTACGGATTCTGG 1108
Db 829 CTTTTTCTGGTGAAGGAGCTCTGGGTCTCTGTGAAGGTGGCCATCTATGGCTTCTGG 888
QY 1109 CTTTTTCTTAAGACTATCGAAGTACCCCACTCAGTCCACCTACTATGATAAATGTTA 1168
Db 889 CTTTTTCTGTGAATATGATGAGCAGCCCATCAGCCACCACTACTATGACAACTCTTA 948
QY 1169 CCTAAGCATGTTTCCACAGATTCCTTAAGAAATACAGCCAAATGCTCCAGTCCATATG 1228
Db 949 CTTTTTCTGGCTTCCATGCCATTCCTCGAGGAATTTCTCCAACTCTGGTATCTTCTATAA 1008
QY 1229 AGAGGAATCTCAAACTGCAATTCAGCAAAATGTGAACGGCTTAAC 1274
Db 1009 ATCGGTGCACTGAGNATGCACTGGACCCATGTGAAGATACCTCAC 1054

RESULT 9

US-10-172-118-859
; Sequence 859, Application US/10172118
; Publication No. US20030224374A1
; GENERAL INFORMATION:
; APPLICANT: Dai, Hongyue
; APPLICANT: He, Yudong
; APPLICANT: Linsley, Peter
; APPLICANT: Mao, Mao
; APPLICANT: Roberts, Chris
; APPLICANT: Van 't Veer, Laura
; APPLICANT: Van de Vijver, Marc
; APPLICANT: Bernards, Rene
; TITLE OF INVENTION: Diagnosis and Prognosis of Breast Cancer Patients
; FILE REFERENCE: 9301-175-999
; CURRENT APPLICATION NUMBER: US/10/172,118
; CURRENT FILING DATE: 2002-06-14
; PRIOR APPLICATION NUMBER: 60/380,170
; PRIOR FILING DATE: 2002-05-14
; NUMBER OF SEQ ID NOS: 2699
; SEQ ID NO 859
; LENGTH: 2117
; TYPE: DNA
; ORGANISM: Homo sapiens
; PUBLICATION INFORMATION:
; DATABASE ACCESSION NUMBER: NM_003034
; DATABASE ENTRY DATE: 2001-06-18
US-10-172-118-859

Query Match 7.1%; Score 224.4; DB 7; Length 2117;
Best Local Similarity 55.3%; Pred. No. 1.8e-46;
Matches 457; Conservative 0; Mismatches 366; Indels 3; Gaps 1;
QY 449 TTTAGACAAACTGGCTTCTGTTGCCATGCCATTCAGACTTCGTGGTTTCCAGAAC 508
Db 714 TTCAGGAAACAATGGAGACTTCGCGACCTGCTCCCATCTCTTGTCTATGACTAAAAATG 773
QY 509 AACACTCCAGTGGGACTTAACATGAGCTACGAGTGGAAAGCAAGAAACACATCCCCATT 568
Db 774 AATTCCTCTATGGGGAAGAGCATTTGGTATGACGGGGAGTTTTTATATCTCATTTACCAT 833
QY 569 CGAGAGAACATTTTCCACATGTTTCCAGTGTGCGAGCCTTTTGTGGACTATCCCTATTAAC 628
Db 834 GACAAATCACTTACTCTCTTCTTCCAGGCAACCCCAT---TCCAGCTGCCATTGAAG 890

QY 629 CAGTGTGAGTGGTGGTAAATGGGGAATTCACAAAGTCTCTCTGCGGAGCAGAAATT 688
 DB 891 AAATCGCGGTGGTGGGAAATGGTGGGATTCGAAGAGAGTGGCTGTGGCGTCAATA 950
 QY 689 GATAAATCTGACTTCCTCTTCAAGTGTAACTCCCTCCCAATCACAGGAGGCGCTAGTAAA 748
 DB 951 GATGAAGCAATTTTGTATGCGATGCAATCTCCCTCTCTTGTCAAGTGAATACACTAAG 1010
 QY 749 GATGTGGAAGCAAAATCTTGTGATGCTGCAATCCAGCAATTAACCTCTGAAGTAC 808
 DB 1011 GATGTGGAATCCAAAGTCAAGTGTGATGCAATCCAGCAATTAACCTCTGAAGTAC 1070
 QY 809 CAGATTTGAAGGAGAAAGACAGTGTGAGGAGCACTCCACCTATGAGATGCA 868
 DB 1071 CAGAACTCTCTGTGTCAGAAAGACATTTGTGGACAAATGAAATCTATAACACAGT 1130
 QY 869 TTCTCTCTCTGCGGAGCAATTTCTATCGGGCAACACAGGAGTCTCTTTAAAGTCTAC 928
 DB 1131 TACATCTACATGCGCTGCTTTCTATGAAGACAGGAAACAGAGCCATCTTTGAGGTTAT 1190
 QY 929 CAACACTCAAGAGTCAAAATAGGCAAAAGTCTCTTCTTCCATCCAGGTAACCTG 988
 DB 1191 TATACACTGTGAGTGTGGTGGCAATCAACAGTGTCTGTGCAACCCCAACTTCTG 1250
 QY 989 AGACACTGCTCTTTCTGGAAGTAAAGGGTGAAGTACCTGCTGTCACAGGC 1048
 DB 1251 CGTAGCATTTGGAAGTCTGGAAGTGTGGAAGTGTGGAAGTGTGGAAGTGTGGAAG 1310
 QY 1049 TTGATGATGCAAGTGTGCTGTGGAAGTGTGGAAGTGTGGAAGTGTGGAAGTGTGGAAG 1108
 DB 1311 CTTTCTCTGAGGAGCTCTGAGTCTGAGTCTGAGTCTGAGTCTGAGTCTGAGTCTGAGT 1370
 QY 1109 CTTTCTCTGAGTCTGAGTCTGAGTCTGAGTCTGAGTCTGAGTCTGAGTCTGAGTCTGAGT 1168
 DB 1371 CTTTCTCTGAGTCTGAGTCTGAGTCTGAGTCTGAGTCTGAGTCTGAGTCTGAGTCTGAGT 1430
 QY 1169 CTTAAGTGTGTTCCAGGAGTCTGAGTCTGAGTCTGAGTCTGAGTCTGAGTCTGAGTCTGAGT 1228
 DB 1431 CTTTCTCTGAGTCTGAGTCTGAGTCTGAGTCTGAGTCTGAGTCTGAGTCTGAGTCTGAGT 1490
 QY 1229 AGAGAACTCTCAAACTGCAATTCAGCAATGTGAAGCGGCTTAAC 1274
 DB 1491 ATCGGTGCACTGAGATGCAATGCACTGCACTGCACTGCACTGCACTGCACTGCACTGCACT 1536

RESULT 10

US-10-085-117-65
 ; Sequence 65, Application US/10085117
 ; Publication No. US2003023334A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Morris, David W.
 ; APPLICANT: Engelhard, Eric K.
 ; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR CANCER
 ; FILE REFERENCE: 529452000121
 ; CURRENT APPLICATION NUMBER: US/10/085,117
 ; PENDING FILING DATE: 2002-02-27
 ; PRIOR APPLICATION NUMBER: US 09/798,586
 ; PRIOR FILING DATE: 2001-03-02
 ; NUMBER OF SEQ ID NOS: 361
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 65
 ; LENGTH: 2117
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; US-10-085-117-65

Query Match 7.1%; Score 224.4; DB 7; Length 2117;
 Best Local Similarity 55.3%; Pred. No. 1.8e-46;
 Matches 457; Conservative 0; Mismatches 366; Indels 3; Gaps 1;

QY 449 TTTAGAGCAAACTGCTTCTGTTGGATGCGATGCACTTCAAGACTCTGTTTCCAGAAC 508

RESULT 11

US-10-430-325-1
 ; Sequence 1, Application US/10430325
 ; Publication No. US20040002138A1
 ; GENERAL INFORMATION:
 ; APPLICANT: SASAKI, KATSUTOSHI
 ; MIURA, KAZUMI
 ; HANAI, NOBUO
 ; NISHI, TATSUNARI
 ; TITLE OF INVENTION: '-2,8-S-ALYLTRANSFERASE
 ; NUMBER OF SEQUENCES: 7
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: NIXON & VANERHYE P.C.
 ; STREET: 1100 NORTH GLENE ROAD
 ; CITY: ARLINGTON
 ; STATE: VIRGINIA
 ; COUNTRY: U.S.A.
 ; ZIP: 22201-4714

DB 714 TTCAGGAAACAAATGGAAGACTCTGCGACCCCTGCCATCTCTTTGCTATGACTAAAAATG 773
 QY 509 AACACTCAGTGGGACTTAACATGAGCTACGAGTGGAAAGCAAGAACACATCCCCATT 568
 DB 774 AATTCCCTATGCGGGAAGAGCAATGGTATGACGCGGAGTTTTTATCTCATTTCCACT 833
 QY 569 CGAGAGAAACATTTTCCACATGTTCCAGTGTGCGAGCCCTTTTGTGGACTATCCCTTATAAC 628
 DB 834 GACAAATCAATTAATCTCTCTTCCAGGCAACCCCAT--TCCAGTCCCATTTGAAG 890
 QY 629 CAGTGTGAGTGGTGGTAAATGCGGAATTTCTCAACAGTCTCTCTGCGGAGCAGAAATT 688
 DB 891 AAATGCGCGTGGTGGGAAATGCGGAATTTCTGAAGAAGAGTGGCTGTGGCGCTCAATA 950
 QY 689 GATAAATCTGACTTCGCTCTTCAAGTGTAACTCCCTCCCAATCACAGGAGGCGCTAGTAAA 748
 DB 951 GATGAAGCAATTTTGTGATGCGAATGCAATCTCCCTCTCTTGTCAAGTGAATACACTAAG 1010
 QY 749 GATGTGGAAGCAAAACAAATCTTGTGACTGTCTAATCCAGCAATTAACCTCTGAAGTAC 808
 DB 1011 GATGTGGAATCCAAAGTCAAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGT 1070
 QY 809 CAGAAATTTGAAGGAGAAAGAGCAAGTGTGAGGAGCACTCTCCACTATGAGATGCA 868
 DB 1071 CAGAACTCTGCTGTGCTCCAGAAACATTTGTGCAACAACTGAAATCTATAACACAGT 1130
 QY 869 TTCTCTCTCTGCGGAGCAATTTCTATCGGGCAACACAGGAGTCTCTTTAAAGTCTAC 928
 DB 1131 TACATCTACATGCGCTGCTTTCTATGAAGACAGGAAACAGAGCCATCTTTGAGGTTAT 1190
 QY 929 CAACACTCAAGAGTCAAAATAGGCAAAAGTCTCTTCTTCCATCCAGGTAACCTG 988
 DB 1191 TATACACTGTGAGTGTGGTGGCAATCAACAGTGTCTGTGCAACCCCAACTTCTG 1250
 QY 989 AGACACTGCTCTTTCTGGAAGTAAAGGGTGAAGTACCTGCTGTCACAGGC 1048
 DB 1251 CGTAGCATTTGGAAGTCTGGAAGTGTGGAAGTGTGGAAGTGTGGAAGTGTGGAAG 1310
 QY 1049 TTGATGATGCAAGTGTGCTGTGGAAGTGTGGAAGTGTGGAAGTGTGGAAGTGTGGAAG 1108
 DB 1311 CTTTCTCTGAGGAGCTCTGAGTCTGAGTCTGAGTCTGAGTCTGAGTCTGAGTCTGAGT 1370
 QY 1109 CTTTCTCTGAGTCTGAGTCTGAGTCTGAGTCTGAGTCTGAGTCTGAGTCTGAGTCTGAGT 1168
 DB 1371 CTTTCTCTGAGTCTGAGTCTGAGTCTGAGTCTGAGTCTGAGTCTGAGTCTGAGTCTGAGT 1430
 QY 1169 CTTAAGTGTGTTCCAGGAGTCTGAGTCTGAGTCTGAGTCTGAGTCTGAGTCTGAGTCTGAGT 1228
 DB 1431 CTTTCTCTGAGTCTGAGTCTGAGTCTGAGTCTGAGTCTGAGTCTGAGTCTGAGTCTGAGT 1490
 QY 1229 AGAGAACTCTCAAACTGCAATTCAGCAATGTGAAGCGGCTTAAC 1274
 DB 1491 ATCGGTGCACTGAGATGCAATGCACTGCACTGCACTGCACTGCACTGCACTGCACTGCACT 1536

```
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
  APPLICATION NUMBER: US/10/430,325
  FILING DATE: 07-May-2003
  CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
  APPLICATION NUMBER: US/08/361,304A
  FILING DATE: 29-NOV-1994
  APPLICATION NUMBER: PCT/JP94/00495
  FILING DATE: 28-MAR-1994
  APPLICATION NUMBER: JP HEI-5-69988
  FILING DATE: 29-MAR-1993
ATTORNEY/AGENT INFORMATION:
  NAME: WILSON, MARY J.
  REGISTRATION NUMBER: 32,955
  REFERENCE/DOCKET NUMBER: 249-66
TELECOMMUNICATION INFORMATION:
  TELEPHONE: (703) 816-4000
  TELEFAX: (703) 816-4100
  TELEX: 200797 NIXN UR
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
  LENGTH: 2117
  TYPE: nucleic acid
  STRANDEDNESS: double
  TOPOLOGY: linear
  MOLECULE TYPE: cDNA to mRNA
  ORIGINAL SOURCE:
    ORGANISM: Homo sapiens
    STRAIN: WM266-4 cell
    CELL TYPE: melanoma
  SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-10-430-325-1

Query Match      7.1%; Score 224.4; DB 7; Length 2117;
Best Local Similarity 55.3%; Pred. No. 1.8e-46;
Matches 457; Conservative 0; Mismatches 366; Indels 3; Gaps 1;

QY 449 TTAGAGCAAACTCGCTTCCTGTTGGGATCCATTCAAGACTTCGTTGGTTCCTCCAGAAC 508
DB 714 TTCAGAAACAAATGGAAGACTCTGCGACCCCTGCCCATCTCTTTGCTATGACTAAATG 773
QY 509 AACACTCCAGTGGGAGCTTAACATGAGCTACGAGGTGGAAGCAAGAAACACATCCCCATT 568
DB 774 AATCCCTATGGGAGAGCATGTGGTATGACGGGGAGTTTATCTACTTCACTTACCATT 833
QY 569 CGAGAGAACATTTTCCACATGTTTCCAGTGTGCGAGCCCTTTTGTGGACTATCCCTATAAC 628
DB 834 GACAATTCACCTTACTCTCTCTTCCACAGGCAACCCCAT---TCCAGCTGCCATTGAAG 890
QY 629 CAGTGTGAGTGGTGGTAAATGAGGGAATTCACAAAGTCTCTCTGCGGAGCAGAAATT 688

QY 449 TTTAGAGCAAACTCGCTTCCTGTTGGGATCCATTCAAGACTTCGTTGGTTCCTCCAGAAC 508
DB 714 TTCAGAAACAAATGGAAGACTCTGCGACCCCTGCCCATCTCTTTGCTATGACTAAATG 773
QY 509 AACACTCCAGTGGGAGCTTAACATGAGCTACGAGGTGGAAGCAAGAAACACATCCCCATT 568
DB 774 AATCCCTATGGGAGAGCATGTGGTATGACGGGGAGTTTATCTACTTCACTTACCATT 833
QY 569 CGAGAGAACATTTTCCACATGTTTCCAGTGTGCGAGCCCTTTTGTGGACTATCCCTATAAC 628
DB 834 GACAATTCACCTTACTCTCTCTTCCACAGGCAACCCCAT---TCCAGCTGCCATTGAAG 890
QY 629 CAGTGTGAGTGGTGGTAAATGAGGGAATTCACAAAGTCTCTCTGCGGAGCAGAAATT 688
```

[illegible]

```

RESULT 13
US-10-085-117-63
; Sequence 63, Application US/10085117
; Publication No. US20030232334A1
; GENERAL INFORMATION:
; APPLICANT: Morris, David W.
; APPLICANT: Engelhard, Eric K.
; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR CANCER
; FILE REFERENCE: 529452000121
; CURRENT APPLICATION NUMBER: US/10/085,117
; CURRENT FILING DATE: 2002-02-27
; PRIOR APPLICATION NUMBER: US 09/798,586
; PRIOR FILING DATE: 2001-03-02
; NUMBER OF SEQ ID NOS: 361
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 63
; LENGTH: 1026
; TYPE: DNA
; ORGANISM: Mus musculus
US-10-085-117-63

```

QY		509	AACACTCCAGTGGGACTAAACAATGAGTTACGAGGTGGAAAGCAAGAACAACATCTCCCATT	568
DB		247	AACTCCCCCATTTGGGAAGAGCCTGTGGTAGATGGGAGAGCTCTATATACTCGTTTCACCATC	306
QY		569	CGAGAGAACATTTTTCCACATGTTTTCCAGTGTGCAGCGCTTTTGCTGGACTATCCCTATAAC	628
DB		307	GATAAATTCACGTA TACTCCCTCTTTTCCCCAGGCAACCCCT---TCCAGCTGCCAATTGAAG	363
QY		629	CAGTGTGCACTGGTGGTAAATGCGGAATTTCTCAAACAAGTCTCTCTGCGGAGCAGAAAATT	688
DB		364	AAATGTGCGGTGGTGGGAACGCTGGGATTTCTGAAGATGAGTGGCTGTGTCCTCGTCAAATA	423
QY		689	GATAAATCTGACTTCGTCTTCAGTGTGAACCTCCCTCCCAATACACAGGAGCGCTAGTAAA	748
DB		424	GATGAACCAAAATTTTGTCAATGCGCTGCAACCTCCCTCTTGTCAAGTGAGTACACCCAGA	483
QY		749	GATGTTGGAGCAAAACAATACTGTGACTGTCAATCCCAAGCATTAATAACCTTGAAGTAC	808
DB		484	GATGTCGGTTCAAAACTCAGTTTGGTGA CAGCTAA CCCCAAGCAATAATTGCCAGAGATTT	543
QY		809	CAGAAATTTGAAGGAGAGAAAGCAAGTTTTTTTGGAGGACATCTCCACCTATGGAGATGCA	868
DB		544	GA AAA CCTGCTGTGTC CGA A A A G A T T T G T G G A C A A C A T G A A G A T C T A T A C C A C A G T	603
QY		869	TTCTCTCTCTGCGCAGCATTTTCTATCGGGCAACACAGGCAATCTCTTTTAAAGTCTAC	928
DB		604	TACATCTACATGCTGCTGCTTTTCTATGAAGACAGGACAGAGCGCTCTCTCCGTGTGTAT	663
QY		929	CAACACTCAAGAGTCAAAAATGAGGCAAAAGGTTCTCTTCTTCCATCCCAGGTACCTG	988
DB		664	TACACACTGAAGATGTTGGAGCAATCAAAACAGTGTCTTTTGTGTAACCCCAACTTTCCT	723
QY		989	AGACACTCTGCTCTTTTCTGGACAACTAAAGGGGTGACTGCATACGGCTTGTCCACAGGC	1048
DB		724	CGNA CATTTGGAAGTCTGMAAAGATAGGGGAA TCACG CCAAGCGCTGTCTACAGGA	783
QY		1049	TTGATGANTGCAAGTGTGCTGTGSAACGTGTGTA AAAAGTGAAGCTCTACGGA TTCTGG	1108
DB		784	CTCTTTTGGTGAGTGCAGCGCTTTGGGCTCTGTGAGGAGGTGCCATCTATGCTTCTGG	843
QY		1109	CCTTTCTCTAAGACTATCGAAGACACCCCACTCAGTCAACCACTACTATGATATACATGTTA	1168
DB		844	CCCTTCTCGGTGAA CATGCAAGCGSACCTTATCAGTCA CCAC TACTATGATCAACAGCTTG	903
QY		1169	CCTAAGCATGGTTTTCCACCAGATSCCTTAAAGAAATACAGCCAAATGCTCCAGCTCCCATATG	1228
DB		904	CCCTTCTCAGGCTACCATGCAATSCCTGTAGGAATTCCTTCAGCTTTGGTATCTTCACAA	963
QY		1229	AGAGGAATCTCAAACTGCAATTTGAGCAATGTGA AACGGCTTAA C	1274
b	b	964	ATCGGCGCGCTGAGGATGCACTGAGACCCGCTGTGAGGAGCGCTCAC	1009

```

RESULT 14
US-10-085-117-62
; Sequence 62, Application US/10085117
; Publication No. US20030232334A1
; GENERAL INFORMATION:
; APPLICANT: Morris, David W.
; APPLICANT: Engelhard, Eric K.
; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR CANCER
; FILE REFERENCE: 529452000121
; CURRENT APPLICATION NUMBER: US/10 085.117
; CURRENT FILING DATE: 2002-02-27
; PRIOR APPLICATION NUMBER: US 09/790,586
; PRIOR FILING DATE: 2001-03-02
; NUMBER OF SEQ ID NOS: 361
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 62
; LENGTH: 1388
; TYPE: DNA
; ORGANISM: Mus musculus
US-10-085-117-62

```

```

Query Match      6.6%; Score 210; DB 7; Length 1388;
Best Local Similarity 54.2%; Prior No. 7.5e-43;
Matches 448; Conservative 0; Mismatches 375; Indels 3; Gaps 1;

QY 449 TTTAGAGCAAACTGGCTTCCTGTTGCGATGCCATTCAAGACTTCGTGTTTCCAGAAC 508
DB 537 TTCAGAGCAAAATGGAAGACTGCTGTGACCTGCCCATCTCTTTGCTATGACTAAGTG 596

QY 509 AACACTCCAGTGGGAGTAACTAGCTACGAGGTGAAAGCAAGAAACACATCCCAATT 568
DB 597 AACTCCCCCATGGGAAGCCCTGTGTAGATGGGAGCTCTTATCTCTGTTCCATC 656

QY 569 CGAGAGCAATTTTCCATGTTTCCAGTGTGCGAGCCTTTTGTGAGCATATCCCTATAC 628
DB 657 GATAATTCACGACTCTCCCTCTTCCCCAGGCAACCCCT---TCCAGTGCATTTGAAG 713

QY 629 CAGTGTGAGTGTGGTAAATGGGGAATTTCTCAACAAGTCTCTCTGCGGAGCAGAAAT 688
DB 714 AATGTGCGGTGGGAAACCGTGGGATTTCTGAAGATGAGTGGCTGTGCCCGTCAAAATA 773

QY 689 GATAAATCTGACTCTGCTCTTCAAGGTAACTCCCTCCCAATCACAGGGAGCGCTAGTAA 748
DB 774 GATGAACCAAAATTTGTCTATGCGGTGCAACTGCTCCCTCTGCTCAAGTGAGTACACCA 833

QY 749 GATGTGGAGCAAAACAAATCTTGTGACTGTCAATCCAGCATATTAACCTGGAAGTAC 808
DB 834 GATGTGCGGTTCAAAACCTCAGTTGGTGACAGCTAACCCCGCAGCAATAATTGCGCAGAGAT 893

QY 809 CAGAAATTTGAGGAGAAAGCAAGTGTGAGGAGCATCTCCACCTATGAGATGCA 868
DB 894 GAAAACCTGCTGTGTGTCAGAAAGAGTTTGGGCAACATGAAGATCTATAACCAACAGT 953

QY 869 TTCCTCTCTGCGCAGCAATTTTCTATCGGGCAACACAGGCGATCTCTTTTAAAGTCTAC 928
DB 954 TACATCTACATGCTGCTCTTTCTATGAAGACAGGCAAGAGCGCTCTCTCGTGTGTAT 1013

QY 929 CAACACTCAAGAGTCAAAATAGAGCAAAAGTTCTTCTTCCATCCCAAGTACCTG 988
DB 1014 TACACACTGAAAGATGTTGGAGCAATCAACAGTGTCTTTTGTCTAAACCCCACTTTCT 1073

QY 989 AGACACTGCTCTTTTCTGGAGAACTAAAGGGTGAAGTACATGCTGCTGTCACAGGC 1048
DB 1074 CGGAACATTTGGAAGTCTGGAAGAGTAGGGGATCCACGCCAAGCGCTGTGCTACAGGA 1133

QY 1049 TTGATGATGCAAGTGTGCTGTGAACTGTGTGAAACGTTGAAGCTCTACGGATTTCTGG 1108
DB 1134 CTCCTTTTGGTGAAGTGTGAGCTTGGGCTCTGTGAGGAGTGTCCATCTATGGCTTCTGG 1193

QY 1109 CCTTCTCTAAGACTATCGAAGACACCCCACTCAGTCAACCTACTATGATAACATGTTA 1168
DB 1194 CCTTCTCGGTGAACATGCAAGGGGACCTATCAGTCAACCACTACTATGACAACTCTTG 1253

QY 1169 CTAAGCATGTTTCCACAGATGCTTAAAGATACAGCAAAATGCTCCAGTCCATATG 1228
DB 1254 CCTTCTCAGGCTACCATGCCATGCTGAGGAATTTCTTCACTTTGTTATCTTCCAAA 1313

QY 1229 AGAGGAATCTCAAACTGCAATTCAGAAATGTGAAACGGCTTAC 1274
DB 1314 ATCGCGCGCTGAGGATGAGCTGAGCCCGTGTGAGGAGCGCTAC 1359

```

RESULT 15

```

US-11-136-527-2140
; Sequence 2140, Application US/11136527
; Publication No. US20050287570A1
; GENERAL INFORMATION:
; APPLICANT: Wyeth
; APPLICANT: Mounts, William M
; TITLE OF INVENTION: Probe Arrays For Expression Profiling of Rat Genes
; FILE REFERENCE: 031896-041000 (AM101086)
; CURRENT APPLICATION NUMBER: US/11/136,527
; CURRENT FILING DATE: 2005-05-25

```

```

; PRIOR APPLICATION NUMBER: US 60/574,294
; PRIOR FILING DATE: 2005-05-26
; NUMBER OF SEQ ID NOS: 362830
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 2140
; LENGTH: 1223
; TYPE: DNA
; ORGANISM: Rattus norvegicus
US-11-136-527-2140

Query Match      6.5%; Score 206.6; DB 16; Length 1223;
Best Local Similarity 53.7%; Prior No. 5.2e-42;
Matches 441; Conservative 4; Mismatches 373; Indels 3; Gaps 1;

QY 449 TTTAGAGCAAACTGGCTTCCTCTTTCGATGCCATTCAAGACTTCGTGTTTCCAGAAC 508
DB 280 TTCAGAGCAAAATGGAAGACTCTGCAACCCGCCCATCTCTTTGCAATGACGAGGTG 339

QY 509 AACACTCCAGTGGGAGCTAAACATGAGCTACGAGGTGGAAGCAAGAAACACATCCCAATT 568
DB 340 AACTCCCCCATGGGGAAGAGCCCTGTGTATGACGGGAGTTTATATCTCGCTCACCATC 399

QY 569 CGAGAGCAATTTTCCACATGTTTCCAGTGTGCGAGCCTTTTGTGAGCTATCCCTATAC 628
DB 400 GACAAATTCACGTAATCTCCCTTTTCCCGAGGCAACCCCT---TCCAGTGCATTTGAAG 456

QY 629 CAGTGTGAGTGGTGTGTAATGCGGGAATTTCTCAACAAAGTCTCTCTGCGGAGCAGAAAT 688
DB 457 AATGTGCGGTGGTGGGAAACGCTGGGATTTCTGAAGATGAGTGGCTGTGGCCGTCAAAATA 516

QY 689 GATAAATCTGACTTCTGCTTCAATGTAACCTTCCCAATCACAGGGAGCGCTAGTAA 748
DB 517 GATGAAGCAAAATTTGTGTCATGCTGTGTAACCTTCTCCCTTGTCAAGTGAATACACAGA 576

QY 749 GATGTGGAAGCAAAACAAATCTGTGACTGTCAATCCAGCATTTATAACCTGAGTAC 808
DB 577 GATGTGGTTCACAGCTCAGTTGAGTACAGTAAACCCAGCATTAATTCGCAGAGATTT 636

QY 809 CAGAAATTTGAAGAGAGAAAGAACAGTGTGAGGAGCATCTCCACCTATGAGATGCA 868
DB 637 GAAAACCTGCTGTGTCCAGAAACAAAGTTTGTGGACACATCAAGATCTATAACCACT 696

QY 869 TTCCTCTCTGCGCAGCATTTTCTATCGGGCAACACAGGATCTCTTTTAAAGTCTAC 928
DB 697 TACATATACATGCTGCTCTTTCTATGAAGACGGGAAACAGAGCCGCTCTCTCGTGTGTAC 756

QY 929 CAACACTCAAGAGTCAAAATGAGGCAAAAGTTCCTCTTCTTCCATCCCAAGTACCTG 988
DB 757 TACACCTGAAAGACGCTGGAGCAATCAAAAGTGTCTTTTGTAAACCCCACTTTCTT 816

QY 989 AGACACTGCTCTCTTTTCTGGAACAACTAAAGGGTGAAGTACATCCGCTTGTCCACAGGC 1048
DB 817 CGAAACATTTGAAAGTCTGGAAGGTTAGGGGATCCATGCCAGGCGCTGTCTACAGGA 876

QY 1049 TTGATGATGCAAGTGTGCTGTGAACTGTGTGAAACGTTGAAGCTCTACGATTTCTGG 1108
DB 877 CTCCTTTTGGTGAAGTGTGAGCCTTGGGCTTTGTGAGGAGTGTCCATCTACGCTTCTGG 936

QY 1109 CCTTCTCTAAGACTATCGAAGAACCCCACTCAGTCAACCTACTATGATAACATGTTA 1168
DB 937 CCTTCTCGGTGAACATGCAAGGCAACCTATACGCCACCATTAATGATGACAACTCTTG 996

QY 1169 CCTAAGCATGTTTCCACAGATGCTTCCCAAGCAATGCTCCAGTCCCATATG 1228
DB 997 CCTTCTCAGGCTTCATGCGCAATGCGGAGGAGTTTCTTCACTCTGCTGTAATCTTCAATA 1056

QY 1229 AGAGGAATCTCAAACTGCAATTTAGCAAAATGTGAAACGGC 1269
DB 1057 ATGGGTGCACTGAGGATGCAACTGGAACCATGTGAGRASCC 1097

```

Search completed: May 31, 2006, 20:43:31
Job time : 3710 secs

GenCore version 5.1.8
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 31, 2006, 19:15:45 ; Search time 50 Seconds
(without alignments)
7457.448 Million cell updates/sec

Title: US-10-501-930-2
Perfect score: 3166
Sequence: 1 cggagcggagtcggtgcc.....gctattagcaaaaaaaaaa 3166

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 246837 seqs, 58886990 residues

Total number of hits satisfying chosen parameters: 493674

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : Published Applications NA New:
1: /EMC Celerra_SIDS3/ptodata/2/pubpna/US09_NEW_PUB.seq.*
2: /EMC Celerra_SIDS3/ptodata/2/pubpna/US06_NEW_PUB.seq.*
3: /EMC Celerra_SIDS3/ptodata/2/pubpna/US07_NEW_PUB.seq.*
4: /EMC Celerra_SIDS3/ptodata/2/pubpna/US08_NEW_PUB.seq.*
5: /EMC Celerra_SIDS3/ptodata/2/pubpna/PCT_NEW_PUB.seq.*
6: /EMC Celerra_SIDS3/ptodata/2/pubpna/US10_NEW_PUB.seq.*
7: /EMC Celerra_SIDS3/ptodata/2/pubpna/US11_NEW_PUB.seq.*
8: /EMC Celerra_SIDS3/ptodata/2/pubpna/US60_NEW_PUB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	42.8	1.4	1556	6	US-10-505-928-322
2	39.8	1.3	541	6	US-10-488-619-949
3	39.8	1.3	641	6	US-10-488-619-1577
4	39.2	1.2	781	6	US-10-953-349-3868
5	39	1.2	1134	7	US-11-217-529-166160
6	38.2	1.2	8391	7	US-11-217-529-166179
7	38.2	1.2	70665	6	US-10-505-928-596
8	38	1.2	3079	7	US-11-293-697-498
9	36.6	1.2	1837	6	US-10-953-349-35182
10	36.2	1.1	1611	7	US-11-217-529-81129
11	36.2	1.1	2296	6	US-10-473-173-24
12	36.2	1.1	37426	6	US-10-473-173-32
13	36	1.1	1241	6	US-10-953-349-22022
14	36	1.1	4908	6	US-10-505-928-226
15	35.8	1.1	510	6	US-10-488-619-1510
16	35.6	1.1	1731	7	US-11-217-529-78208
17	35.4	1.1	785	6	US-10-946-650-15
18	35.4	1.1	2037	7	US-11-266-446-75
19	35.4	1.1	3844	6	US-10-946-650-32
20	35.2	1.1	2480	7	US-11-293-697-1939
21	35	1.1	669	7	US-11-233-726-37
22	35	1.1	1563	7	US-11-217-529-82562
23	35	1.1	2103	7	US-11-293-697-69
24	34.8	1.1	300	6	US-10-488-619-2492
25	34.8	1.1	687	7	US-11-242-317-38

26	34.8	1.1	2983	7	US-11-293-697-589	Sequence 589, App
27	34.8	1.1	3288	6	US-10-196-749-477	Sequence 477, App
28	34.6	1.1	847	6	US-10-525-126-99	Sequence 99, Appl
29	34.6	1.1	1887	7	US-11-217-529-4458	Sequence 4458, Ap
30	34.4	1.1	1578	7	US-11-217-529-76837	Sequence 76837, A
31	34.4	1.1	5853	7	US-11-217-529-1030	Sequence 1030, Ap
32	34.2	1.1	1997	7	US-11-293-697-546	Sequence 546, App
33	34.2	1.1	2297	7	US-11-266-446-62	Sequence 62, Appl
34	34.2	1.1	2476	7	US-11-217-529-166168	Sequence 166168,
35	34.2	1.1	4941	6	US-10-713-648A-34	Sequence 34, Appl
36	34	1.1	600	7	US-11-217-529-166170	Sequence 166170,
37	34	1.1	3460	7	US-11-293-697-1714	Sequence 1714, Ap
38	33.8	1.1	958	6	US-10-505-928-315	Sequence 315, App
39	33.8	1.1	1185	6	US-10-953-349-14579	Sequence 14579, A
40	33.8	1.1	1599	6	US-10-953-349-27865	Sequence 27865, A
41	33.8	1.1	1990	7	US-11-266-446-76	Sequence 76, Appl
42	33.6	1.1	2876	6	US-10-505-928-468	Sequence 468, App
43	33.6	1.1	2884	7	US-11-293-697-700	Sequence 700, App
44	33.4	1.1	600	7	US-11-217-529-166176	Sequence 166176,
45	33.4	1.1	995	6	US-10-953-349-16178	Sequence 16178, A

ALIGNMENTS

RESULT 1
US-10-505-928-322
; Sequence 322, Application US/10505928
; Publication No. US20060088532A1
; GENERAL INFORMATION:
; APPLICANT: Ludwig Institute for Cancer Research et al.
; TITLE OF INVENTION: LYMPHATIC ENDOTHELIAL GENES
; FILE REFERENCE: 28967/39178
; CURRENT APPLICATION NUMBER: US/10/505,928
; PRIOR FILING DATE: 2004-08-27
; PRIOR APPLICATION NUMBER: US 60/369,019
; PRIOR FILING DATE: 2002-03-07
; NUMBER OF SEQ ID NOS: 866
; SOFTWARE: Patentin 3.2
; SEQ ID NO 322
; LENGTH: 1556
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-505-928-322

Query Match	1.4%	Score 42.8;	DB 6;	Length 1556;
Best Local Similarity	54.4%	Pred. No. 0.069;	Mismatches 72;	Indels 0;
Matches	86;	Conservative 0;	Gaps	0;
Qy	608	TTTGTGACATATCCCTATAACCAATGTGCACTGGTGTGTAATGGGGGAATTCACCAAG	667	
Db	630	TTTGACCAACATACCCTGTAAATAATGTGTGGTGGTGTGTAATGGAGGAGTTTGAAGAAT	689	
Qy	668	TCTCTTCGGGAGCAGAAATGTGTAATCTCACTTCGTTTCAGTGTAACTCCCCCA	727	
Db	690	AAGACATTAGGAGAGAAATAATCGATTCCTATGATGTATAATAAAGTAATGAATGGTCTT	749	
Qy	728	ATCACAGGAGCGCTAGTAAAGTGTGGAAGCAAAAC	765	
Db	750	GTTTATGACATGAGAGAGAGTGGGAGAGGACAC	787	

RESULT 2
US-10-488-619-949/c
; Sequence 949, Application US/10488619
; Publication No. US20060099578A1
; GENERAL INFORMATION:
; APPLICANT: Greenlee, Winner and Sullivan, P.C.
; TITLE OF INVENTION: Human Mitochondrial DNA Polymorphisms, Haplogroups, Associations
; FILE REFERENCE: 98-01 WO
; CURRENT APPLICATION NUMBER: US/10/488,619
; CURRENT FILING DATE: 2004-03-01


```

; APPLICANT: ALEXANDROV, Nickolai et al.
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES
; TITLE OF INVENTION: ENCODED THERBY
; FILE REFERENCE: 2750-1579PUS2
; CURRENT APPLICATION NUMBER: US/10/953,349
; CURRENT FILING DATE: 2004-09-30
; NUMBER OF SEQ ID NOS: 40252
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 35182
; LENGTH: 1837
; TYPE: DNA
; ORGANISM: Zea mays subsp. mays
US-10-953-349-35182

```

	Query Match	1.2%;	Score 36.6;	DB 6;	Length 1837;
	Best Local Similarity	53.1%;	Prod. No. 3.4;		
	Matches	78;	Conservative	0; Mismatches 69;	Indels 0; Gaps 0
Qy	2	GGAGCGGGAGTCGGTGC	CGCCCGGGCTCGGCTTC	CGCCCGGACGTTTGGCGGCGAGGA	61
Db	156	GGATTGGCGGCTGAGG	CAGCAGGCGAGCGTGC	CGCCCGGCGGCGGCGCGTG	215
Qy	62	CGCCCGTGGCTCAGGAT	CAGATCGGGGGGCGAGCT	TGTTCGCCCTCATAGG	CAGCCTGATG 121
Db	216	GCCAGCGGCTCGGCG	GTGCGATCGCTCGGCG	CGCGGAGCTCGAGG	CAGCGGTGG 275
Qy	122	CTGCTGCTCTCTCTCG	GCTATGCTCTGG		148
Db	276	CGGCTTCGCGGCGCG	CGCGGCGCGG		302

```

RESULT 10
US-11-217-529-81129/c
; Sequence 81129, Application US/11217529
; Publication No. US2006099612A1
; GENERAL INFORMATION:
; APPLICANT: SUNTORY LIMITED
; APPLICANT: NAKAO, YOSHIHIRO
; APPLICANT: NAKAMURA, NORIHISA
; APPLICANT: KODAMA, YUKIKO
; APPLICANT: FUJIMURA, TOMOKO
; APPLICANT: ASHIKARI, TOSHIHIKO
; TITLE OF INVENTION: METHODS FOR ANALYZING GENES OF INDUSTRIAL YEASTS
; FILE REFERENCE: S-38-285
; CURRENT APPLICATION NUMBER: US/11/217,529
; CURRENT FILING DATE: 2005-09-02
; PRIOR APPLICATION NUMBER: US 10/932,182
; PRIOR FILING DATE: 2004-09-02
; NUMBER OF SEQ ID NOS: 197023
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 81129
; LENGTH: 1611
; TYPE: DNA
; ORGANISM: Saccharomyces pastorianus
; FEATURE:
; NAME/KEY: modified base
; LOCATION: (16)..(16)
; OTHER INFORMATION: a, c, g, t, unknown, or other
US-11-217-529-81129

```

	Query Match	1.1%;	Score 36.2;	DB 7;	Length 1611;
	Best Local Similarity	47.9%;	Pred. No. 4;		
	Matches 104;	Conservative 0;	Mismatches 113;	Indels 0;	Gaps 0;
Qy	2898	ATTTAAACATTTGTAATCTGTTTCAAGTTTACATCTCTTTTCATTCTTTTATAGCAATCA	2957		
Db	494	ATTTTCAAATGCAAACTTTTGTAAGAAGCTAGCACTGTTTAAAGGTTTCAAATCTTCGCA	435		
Qy	2958	AAGCTATTAGTCTCAGAAAAATTTATCAGAAGTTTCATATATATAATTTTTCAGAAAGGGTAAA	3017		
Db	434	TCAGGTTCCAGTGAAGAGTGTGTTAGGAAGAAAGATAATTTTTTTTGTCTGTCTAGTAATAGAGGCTCTTT	375		
Qy	3018	AGGCTTTTTTTGTTAAATAAAATAAAATTTATTTATTTTCTTCTGATGAATAGAGGCTCTTT	3077		

Db 374 AAGTTTCTCTTCGTTATCATATATCTTCTGGGATTTCTTCTATGGAATTCATGATCTCC 315

Qy 3078 TATGCTGCTGTAATGACCTAATTTAGCTTTAAATTA 3114

Db 314 TCTATTGGTTTCATAAAAATTTGAGAGCTTCTAATCA 278

```

RESULT 11
US-10-473-173-24/c
; Sequence 24, Application US/10473173
; Publication No. US2006008823A1
; GENERAL INFORMATION:
; APPLICANT: VAN ANDEL INSTITUTE
; TITLE OF INVENTION: Microarray Gene Expression Profiling in Clear Cell Renal Cell
; TITLE OF INVENTION: Carcinoma: Prognosis and Drug Target Identification
; FILE REFERENCE: 38345-170094
; CURRENT APPLICATION NUMBER: US/10-473,173
; CURRENT FILING DATE: 2003-09-29
; PRIOR APPLICATION NUMBER: US 60/273,411
; PRIOR FILING DATE: 2001-03-29
; NUMBER OF SEQ ID NOS: 498
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 24
; LENGTH: 2296
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-473-173-24

```

	Query Match	1.1%; Score 36.2; DB 6; Length 2296;
	Best Local Similarity 50.9%; Pres. No. 4.9;	
Matches	86; Conservative 0; Mismatches 83; Indels 0; Gaps 0	
Qy	2 GGAGCGCGAGTCGGTCCGCCCGGGCTCGGCTTCGCCCGGCACGCTTTGCGCGCGAGA 61	
Dd	1036 GGAAGGAGCGGGGTGCCGGGCGCGGGGCGCTCCTTTGTGCGCGGGGCTTGCCCGG 977	
Qy	62 CGCCCGTGGGTCCAGGATGAGATCGGGGGGCACGCTGTTCGCCCTCATAGGCAGCGCTGATG 121	
Dd	976 CTGCGGGGGTCGACCCCCCTCGGTGCTCGGGGCGCGCGCGCGCGCGCTACGCCG 917	
Qy	122 CTGCTGCTCCTCTCGGTATGCTATTGCTGCCCAAGCCGACGCGCTGCCC 170	
Dd	916 CCGCTCTCTGGCTTCGGCTTGTCTCGGCGCTTCGCCCGGCAGCTTCGCC 868	

```

RESULT 12
US-10-473-173-32/c
; Sequence 32, Application US/10473173
; Publication No. US20060088823A1
; GENERAL INFORMATION:
; APPLICANT: VAN ANDEL INSTITUTE
; TITLE OF INVENTION: Microarray Gene Expression Profiling in Clear Cell Renal Cell
; TITLE OF INVENTION: Carcinoma: Prognosis and Drug Target Identification
; FILE REFERENCE: 38345-170094
; CURRENT APPLICATION NUMBER: US/10473,173
; CURRENT FILING DATE: 2003-09-29
; PRIOR APPLICATION NUMBER: US 60/219,411
; PRIOR FILING DATE: 2001-03-29
; NUMBER OF SEQ ID NOS: 498
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 32
; LENGTH: 37426
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-473-173-32

```

Query Match	1.1%	Score 36.2;	DB 6;	Length 37426;
Best Local Similarity	47.9%;	Prod. No. 24;		
Matches 104;	Conservative 0;	Mismatches 113;	Indels 0;	Gaps 0;
Qy	2888	GTCACTTGATTTATACATTTGTGATGTTTTCAGTTTACATCTCTTCATTCCTTTA	2947	

Db 18605 GGCACAGATGAATTAACATTTACTGAGCATTATAGATTTTTTTTTTAATGAGTGAATAT 18546
 QY 2948 TAGCAATCAACGATATTAGCTTCAGAAATTTATCAGAGTTTCATATATAATATTTTGC 3007
 Db 18545 AAGGATATGCCAAATTTTAATTTTAAATGAATCAATATTTCTTACTATATTTTCTTT 18486
 QY 3008 AAAGGTTAAAGGCTTTTTTGTGTAATAAATAAATTTATTTTCTCTGATGAATA 3067
 Db 18485 CAAAGTACTTCAAAATTTCTCTTTTAAATTAATTAATTTCTTCAATGTTAAGGAAT 18426
 QY 3068 GAGGCTCTTTATGCTGCTGAATGAACCTAATTAG 3104
 Db 18425 GAAGTAGATCATTTATCTTTTAAAGAGTTGTAATCAG 18389

RESULT 13
 US-10-953-349-22022
 ; Sequence 22022, Application US/10953349
 ; Publication No. US20060107345A1
 ; GENERAL INFORMATION:
 ; APPLICANT: ALEXANDROV, Nikolai et al.
 ; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES
 ; FILE OF INVENTION: ENCODED THERBY
 ; FILE REFERENCE: 2750-1579PUS2
 ; CURRENT APPLICATION NUMBER: US/10/953,349
 ; CURRENT FILING DATE: 2004-09-30
 ; NUMBER OF SEQ ID NOS: 40252
 ; SOFTWARE: PatentIn version 3.3
 ; SEQ ID NO 22022
 ; LENGTH: 1241
 ; TYPE: DNA
 ; ORGANISM: Glycine max
 US-10-953-349-22022

Query Match 1.1%; Score 36; DB 6; Length 1241;
 Best Local Similarity 51.9%; Pred. No. 3.9;
 Matches 81; Conservative 0; Mismatches 75; Indels 0; Gaps 0;
 QY 2898 ATTACATTTTGTATGTTTCAAGTTTACATCTTTTCATCTTTTATAGCAATCA 2957
 Db 351 ATTAATATTTATGATAATTTTATGATTATATAAATTTATTTTATTTTATGCTTTC 410
 QY 2958 AACGTATTAGCTTCAGAAATTTATCAGAGTTTCATATATAATTTTGCAGAGGTTAA 3017
 Db 411 AATTATATTTTATTTAAACAGTACATATATTCAAGTAAATAATTTCAATCATGGCAA 470
 QY 3018 AGGCTTTTTTGTGTAATAAATAAATTTATTTT 3053
 Db 471 AGAAAAATATATTTTATAGATATAATTTGCTTATT 506

RESULT 14
 US-10-505-928-226
 ; Sequence 226, Application US/10505928
 ; Publication No. US20060088532A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Ludwig Institute for Cancer Research et al.
 ; TITLE OF INVENTION: LYMPHATIC ENDOTHELIAL GENES
 ; FILE REFERENCE: 28967/39178
 ; CURRENT APPLICATION NUMBER: US/10/505,928
 ; CURRENT FILING DATE: 2004-08-27
 ; PRIOR APPLICATION NUMBER: US 60/363,019
 ; PRIOR FILING DATE: 2002-03-07
 ; NUMBER OF SEQ ID NOS: 866
 ; SOFTWARE: PatentIn 3.2
 ; SEQ ID NO 226
 ; LENGTH: 4908
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 US-10-505-928-226

Query Match 1.1%; Score 36; DB 6; Length 4908;
 Best Local Similarity 60.2%; Pred. No. 8.5;

Matches 77; Conservative 0; Mismatches 50; Indels 1; Gaps 1;
 QY 2881 ATACAGTGTCACTTGTATTTATTAATTTTGTATGTTTCAAGTTTACATCT-CTTTCA 2939
 Db 3791 ATTGGGTTTCATTTTGGGATTTTGCATGTAATACGTAATTTCTAGTTTTCATATAAAGTAG 3850
 QY 2940 TCTTTTATAGCAAAATCAACGTTATAGCTTCAGAAATTTATCAGAGTTTCATATATAA 2999
 Db 3851 TCTTTTATACAAATGAAGATTTTCTTGATATATTTAATTAATGAATATATAAG 3910
 QY 3000 TATTTTGC 3007
 Db 3911 AACTGTAC 3918

RESULT 15
 US-10-488-619-1510
 ; Sequence 1510, Application US/10488619
 ; Publication No. US20060099578A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Greenlee, Winner and Sullivan, P.C.
 ; TITLE OF INVENTION: Human Mitochondrial DNA Polymorphisms, Haplogroups, Associations
 ; FILE OF INVENTION: Physiological Conditions, And Genotyping Arrays
 ; FILE REFERENCE: 98-01 WO
 ; CURRENT APPLICATION NUMBER: US/10/488,619
 ; CURRENT FILING DATE: 2004-03-01
 ; NUMBER OF SEQ ID NOS: 3040
 ; SOFTWARE: PatentIn version 3.1
 ; SEQ ID NO 1510
 ; LENGTH: 510
 ; TYPE: DNA
 ; ORGANISM: Mus musculus
 US-10-488-619-1510

Query Match 1.1%; Score 35.8; DB 6; Length 510;
 Best Local Similarity 52.3%; Pred. No. 2.6;
 Matches 79; Conservative 0; Mismatches 72; Indels 0; Gaps 0;
 QY 2897 TATTTACATTTGTAAATGTTTTCATCAAGTTTACATCTTTTCATCTTTTATAGCAATC 2956
 Db 259 TCTTAAACTGTTTAAATTTTCTTAAACATTTCTATAGTGAAGGCTATGAATTTCAACC 318
 QY 2957 AAACGTATTAGCTTCAGAAATTTATCAGAGTTTCATATATAATTTTGCAGAGGTTAA 3016
 Db 319 ATAGAAATTTCTCAATTAACAGTTCAAGTATTTCTTATTTAGGATATTTGAAATTTCTAA 378
 QY 3017 AAGCTTTTTTGTGTAATAAATAAATTTA 3047
 Db 379 AACGTTTTTACTTAATTAATTTGTACTGA 409

Search completed: May 31, 2006, 19:18 56
 Job time : 51 secs


```

; NAME: WILSON, MARY J.
; REGISTRATION NUMBER: 32,955
; REFERENCE/DOCKET NUMBER: 249-46
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 816-4000
; TELEFAX: (703) 816-4100
; TELEX: 200797 NIXN UR
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2117
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA to mRNA
; US-08-361-304A-1

Query Match
Best Local Similarity 55.3%; Pred. No. 3.2e-50;
Matches 457; Conservative 0; Mismatches 366; Indels 3; Gaps 1;

QY 449 TTTAGAGAAAACCTGGCTTCCTGTTGGATGCCAATTCAAGACTTCGTGGTTTCCAGAAC 508
Db 714 TTCAGGAAAACAAATGGAAGACTGTTGGACCCCTGCCATCTCTTTGCTATGACTAAAAATG 773
QY 509 AACACTCCAGTGGGACTTAACATTAAGCTACGAGTGGGAACGAACAAACACATCCCCATT 568
Db 774 AATTCCCTCTATGGGGAAGAGCATTTGGTATGACGGGAGTTTTTATATCTCATTTACCATT 833
QY 569 CGAGAGAACATTTTCCACATGTTTCCAGTGTGCGAGCCTTTTGTGGACTATCCCTTATAAC 628
Db 834 GACAAATCAACTTACTCTCTTCCACAGCAACCCCAT---TCCAGTGCCTTGAAG 890
QY 629 CAGTGTGAGTGGTGTGTAATGAGGAAATTTCTCAACAGTCTCTCTGCGGAGCAGAAATT 688
Db 891 AAATGCGGTTGGTGGGAAATGTTGGGATTTCTGAAGAAGAGTGGCTGTGGCGCTCAATA 950
QY 689 GATAAATCTGACTTCGCTTCTCAAGTGTAACTCCGCCCAATCAGAGGAGCGCTAGTAA 748
Db 951 GATGAAGCAAAATTTTGTCAATGCAATCTCCCTCTCTTGTCAAGTGAATACACTAAG 1010
QY 749 GATGTTGGAAGCAAAACAAATCTTGTGACTGTCAATCCAGCAATTAATAACCTCGAAGTAC 808
Db 1011 GATGTTGATCCAAAGTCAAGTGTGATGACAGCTAATCCAGCAATTAATTCGGCAAGTTT 1070
QY 809 CAGAAATTTGAAGGAGAAAGAAAGCACTTTTGGAGGACATCTCCACTATGGAGATGCA 868
Db 1071 CAGAACTTCTGTGTTCCAGAAACACATTTTGTGGAACAACATGAAATCTATAACCCACAGT 1130
QY 869 TTCTCTCTCTCCAGCAATTTCTATCGGGCCCAACACAGGCATCTCTTTTAAAGTCTAC 928
Db 1131 TACATCTACATGCTGCTGCTTTGATGAAGACAGAACAGAGCCATCTTTGAGGTTTAT 1190
QY 929 CAAACACTCAAGAGTCAAAAATTAAGGCAAAAGTTCTCTTCTTCCATCCAGGTACCTG 988
Db 1191 TATACACTGTCAAGTGTGGTGCAATCAACAGTGTCTTTTGGCAACCCCACTTCTG 1250
QY 989 AGACACTCTGCTCTTTTCTGAGAACTAAAGGGGTGACTGATACCGCTTGTCCACAGC 1048
Db 1251 CGTAGCATTTGGAAGTTCTGGAAAGTAGAGGAATCCATGCCAAGCGCTGTCCACAGGA 1310
QY 1049 TTGATGATTCGAAGTGTGCTGTGAACTGTGGAACCGTGAAGCTCTACGGATTCGG 1108
Db 1311 CTTTCTCTGAGCGGAGCTCTGGTCTCTGGAAGAGGTGGCCATCTATGGCTTCTG 1370
QY 1109 CTTTCTCTAAGACTATCGAAGTACCCCACTCAGTCAACCACTACTATGATAAATGTTA 1168
Db 1371 CCCTTCTCTGATATGATGATGACAGCCCACTCAGCCCACTACTATGACACGCTTA 1430
QY 1169 CCTAAGCATGTTTCCACAGATTCCTAAAGAAATACAGCCAAATGCTCCAGCTCATATG 1228
Db 1431 CCCTTTTCTGGCTTCCATGCAATGCCATGCCGAGGAATTTCTCCAACTCTGGTATCTTCAATA 1490
QY 1229 AGAGGAATCTCAAACTGCAATTAAGCAAAATGGAAGCGGCTTAAC 1274

```

RESULT 2

```

; Sequence 1, Application US/08361304A
; Patent No. 6596523
; GENERAL INFORMATION:
; APPLICANT: SASAKI, KATSUTOSHI
; APPLICANT: MIURA, KAZUMI
; APPLICANT: HANAI, NOBUO
; APPLICANT: NISHI, TATSUNARI
; TITLE OF INVENTION: '-2,8-SIALYLTRANSFERASE
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NIXON & VANDERHYE P.C.
; STREET: 1100 NORTH GLEBE ROAD
; CITY: ARLINGTON
; STATE: VIRGINIA
; COUNTRY: U.S.A.
; ZIP: 22201-4714
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/361,304A
; FILING DATE: 29-NOV-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/JP94/00495
; FILING DATE: 28-MAR-1994
; APPLICATION NUMBER: JP HEI-5-69988
; FILING DATE: 29-MAR-1993
; ATTORNEY/AGENT INFORMATION:

```

Db 1491 ATCGGTGCACTGAGATGCACTGGACCCATGTGAATACCTCAC 1536

RESULT 3

US-09-949-016-436
; Sequence 436, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 436
; LENGTH: 2117
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-436

Query Match 7.1%; Score 224.4; DB 3; Length 2117;
Best Local Similarity 55.3%; Pred. No. 3.2e-50;
Matches 457; Conservative 0; Mismatches 366; Indels 3; Gaps 1;
QY 449 TTAGAGCAAACTGGCTTCTGTTGCGATGCCATTTCAAGACTTCGTGTTTCCAGAAC 508
Db 714 TTCAGAAACAATGGAAGACTGCTGGACCTGCCCATCTCTTGTCTATGACTAAATG 773
QY 509 AACCTCCAGTGGGACTAACATGATGATGAGTGGGAAAGCAAGAACACATCCCAT 568
Db 774 AATCCCCCTATGGGGAAGCATGTGGTATGACGGGGAGTTTTTATCTCATTCACCAT 833
QY 569 CGAGAGCAATTTCCACATGTTTCCAGTGTCCGAGCTTTTGTGACTATCCCTATAC 628
Db 834 GACATTCACCTTACTCTCTTCCACAGGCAACCCCAT---TCAGCTGCCATGAG 890
QY 629 CAGTGTGAGTGGTGGTAATGGGGAAATTTCTCAACAGTCTCTCTGCGGAGCAGAAAT 688
Db 891 AAATCGCGGTGGTGGGAAATGTTGGATTTCTGAAGAGAGTGGCTGTGGCGTCAATA 950
QY 689 GATAAATCTGACTTCTGTTTCAAGTGTAACTTCCCGCCCAATCAGAGGAGCGCTAGTAA 748
Db 951 GATGAAGCAAAATTTTGTCTATGCGATGCAATCTCCCTCTCTGTCAGTGAATACATA 1010
QY 749 GATGTTGAAGCAAAACAAATCTCTGACTGTCAATCCAGCATTTAATACCTGAAGTAC 808
Db 1011 GATGTTGATCCAAAGTCACTGTAGTACAGCTAATCCAGCATTAATTCGGAAGGTTT 1070
QY 809 CAGAAATTTGAAGGAGAAAGACACAGTTTTTGGAGGACATCTCCACCTATGGAGATCA 868
Db 1071 CAGAACCTTCTGTGTCAGAAAGACATTTGGGACACATGAATAATCTATAACACAGT 1130
QY 869 TTCCTCTCTGCGCAGCAATTTTCTATCGGGCCAAACAGAGCATCTCTTTAAAGTCTAC 928
Db 1131 TACATCTACATGCTGCTGCTTTCTATGAAGACAGGACAGAGCCATCTTTGAGGGTTAT 1190
QY 929 CAACACTCAAGAGTCAAAATGAGGCAAAAGTTCTTCTTCTCCATCCAGGTACCTG 988
Db 1191 TATACACTGTGAGATGTTGGTGCCCAATCAACAGTGTGTTGGCAACCCCAACTTTCTG 1250
QY 989 AGACACCTCGCTCTTTCTGGAGAACTAAAGGGGTGACTGCATACCGCTTGTCCACAGGC 1048
Db 1251 CGTAGCATTGAAGATTTCTGGAAGAGTAGAGAAATCCATGCCAAGCGCTGTGCCACAGGA 1310

QY 1049 TTGATGATTTGCAAGTGTGCTGTGGAACCTGTGTGGAACCTGTGTACGGAATCTGG 1108
Db 1311 CTTTTTCTGTTGAGCGCAGCTCTGGTCTCTGTGAAGAGGTGCCATCTATGCTTCTGG 1370
QY 1109 CTTTTTCTTAAGACTATCGAAGTACCCCACTCAGTCACCACTACTATGATAACATCTTA 1168
Db 1371 CCTTCTCTGTGAATATGATGATGACGCCATCAGCCACCACTACTATGATAACATCTTA 1430
QY 1169 CTTAAGCATGGTTTCCACCAGATTCCTAAAGAAATACAGCAAAATGCTCCAGCTCCATATG 1228
Db 1431 CCTTTTCTGCTTCCATGCCATGCCGAGGAATTTCTCCAACCTCTGGTATCTTCATAAA 1490
QY 1229 AGAGGAATCTCAAACTGCAATTTAGCAAAATGTGAACCGCTTAAAC 1274
Db 1491 ATCGGTGCACTGAGAATGCACTGACCCATGTGAATACCTCAC 1536

RESULT 4

US-09-949-016-13872
; Sequence 13872, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13872
; LENGTH: 137394
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(137394)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-13872

Query Match 4.4%; Score 139.2; DB 3; Length 137394;
Best Local Similarity 55.9%; Pred. No. 4.4e-26;
Matches 264; Conservative 0; Mismatches 208; Indels 0; Gaps 0;
QY 803 AAGTACCAGAAATTTGAAGGAGAAATAAGCACAGATTTTTTGGAGGACATCTCCACCTATGGA 862
Db 134353 AGGTTTCAGAACCTTCTGTGTCAGAAAGACATTTTGGAGCAACATGAAATCTATAAC 134412
QY 863 GATGATTCCTCTCTGTCGACAGCTTTTTCCTATCGGGCCAAACACAGGACATCTCTTTTAA 922
Db 134413 CACAGTTACATCTACATGCTGCTGCTTTTCTATGAAGACAGAAACAGAGCATCTTTGAGG 134472
QY 923 GTCTTACCACACCTCAAGAGTCAAAATGAGGCAAAAGTTCTCTTCTTCATCCACCG 982
Db 134473 GTTTATTATACACTGTGATGATGTTGGTCCCAATCAACAGTGTCTTTTGGCCAAACCCCAAC 134532
QY 983 TACCTGAGACACCTCGCTCTTTTATGGAGAACTAAAGGGGTGACTGTCATACCGCTTGTCC 1042
Db 134533 TTTCTGGTAGCATTTGGAAGTTTGGAAAAGTAGAGAAATCCATGCCAAGCGCTGTCC 134592
QY 1043 ACAGGCTTGATGATGCAAGTGTGCTGTGGAACCTGTGTGAAAACGTGAAGCTCTACCGA 1102
Db 134593 ACAGGACTTTTCTGTTGAGGCGCTCTGTTGAGAGGTGGCCATCTATGCGC 134652
QY 1103 TTTCTGGCTTTCTCTTAAGACTATGGAAGACACCCCACTCAGTCACCACTACTATGATTAAC 1162
Db 134653 TTTCTGGCCCTTCTCTGTGTAATATCATGAGCAGGCCCATCAGCCCACTACTATGACAAAC 134712


```
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 72792
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-72792

Query Match
Best Local Similarity 3.3%; Score 105.6; DB 3; Length 601;
Matches 198; Conservative 1; Mismatches 155; Indels 0; Gaps 0;

QY 803 AAGTACCAGAAATTGAAGGAGAGAAAGACACAGTTTTTTCGAGGACATCTCCACCTATGGA 862
DB 354 AGTTTCAGAACCTTCTGTGGTCCAGAAAGACATTTGTGGCAACATGAAATYTTATAAC 295
QY 863 GATGCAATCTCTCTCTGTCAGCATTTTCTATCGGGCCCAACACAGGCACTCTCTTTTAA 922
DB 294 CACAGTTACATCTACATGCTGCTTTTCTATGAAGACAGAAACAGCCATCTTTGAGG 235
QY 923 GTCTACCAAACTCAAGAGCTCAAAATAGAGGCAAAAGTTCTCTTCTTCATCCCAAGG 982
DB 234 GTTTATTATACACTGTCAATGTGTGTCATCAATCAACAGTGTCTTTTCCCAACCCCAAC 175
QY 983 TACCTGAGACACTCGCTTTTCTTGGAGAACTAAAGGGTGACTGCATACCGCTTGTC 1042
DB 174 TTTCTGCTAGCATTTGGAAAGTTCTGGAAAGTAGAGGAATCCATGCCAAGCGCTGTCC 115
QY 1043 ACAGCTTTGATGTCAGTGTGCTGTGGAACTGTGTGAAACGTGAAGCTCTACCGA 1102
DB 114 ACAGGACTTTTCTGTGTGAGCGCAGCTCTGGGTCTCTGTGAGAGGTGGCCATCTATGGC 55
QY 1103 TTTGGCTTTCTTAAAGACTATCGAAGACACCCCACTCAGTCACCACTACTAT 1156
DB 54 TTTGGCCCTTCTCTGTGAATATGATGACGAGCCCATCAGCCACCACTACTAT 1

RESULT 8
US-09-949-016-26414/c
; Sequence 26414, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 26414
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-26414

Query Match
Best Local Similarity 3.2%; Score 100.2; DB 3; Length 601;
Matches 162; Conservative 0; Mismatches 103; Indels 0; Gaps 0;

QY 1010 AGAAGTAAAGGGTGACTGCATACCGTTGTCCAGAGCTTGATGATGCAAGTGTGCT 1069
DB 601 AAAAGTAGAGGAATCCATGCGCTGTCCAGAGCTTTTCTGTGAGCGCAGCT 542
QY 1070 GTGGAACCTGTGTGAAACGCTGAACTCTACGATCTGGCTTTCTTAAGACTATCGNA 1129
DB 541 CTGGGTCTCTGTGAAAGAGGTGGCATCTATGATGCTTCTGTGAATATGCAT 482
QY 1130 GACACCCCACTCAGTCACCACTATATGATAACATGTTTACCTAAGCATGTTTCCACCAG 1189
DB 481 GAGCAGCCCATCAGCCACCACTATATGACAAAGCTTTTACCTTTTCTGGCTTCCATGCC 422
QY 1190 ATGCTTAAAGAAATACAGCCAAATCTCCAGTCCATATGAGAGGAATCTCAAACTGCAA 1249
DB 421 ATGCCCAGGAAATTTCTCCAACTTTGGTATCTTCTATAAAATCGGTGCACTGAGAATGCAG 362
QY 1250 TTCAGCAAAATGTGAACCGCTTAAAC 1274
DB 361 CTGGACCCCATGTGAAGATACCTTAAAC 337

RESULT 9
US-09-949-016-72793/c
; Sequence 72793, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 72793
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-72793

Query Match
Best Local Similarity 3.2%; Score 100.2; DB 3; Length 601;
Matches 162; Conservative 0; Mismatches 103; Indels 0; Gaps 0;

QY 1010 AGAAGTAAAGGGTGACTGCATACCGTTGTCCAGAGCTTGATGATGCAAGTGTGCT 1069
DB 601 AAAAGTAGAGGAATCCATGCGCTGTCCAGAGCTTTTCTGTGAGCGCAGCT 542
QY 1070 GTGGAACCTGTGTGAAACGCTGAACTCTACGATCTGGCTTTCTTAAGACTATCGNA 1129
DB 541 CTGGGTCTCTGTGAAAGAGGTGGCATCTATGATGCTTCTGTGAATATGCAT 482
QY 1130 GACACCCCACTCAGTCACCACTATATGATAACATGTTTACCTAAGCATGTTTCCACCAG 1189
DB 481 GAGCAGCCCATCAGCCACCACTATATGACAAAGCTTTTACCTTTTCTGGCTTCCATGCC 422
QY 1190 ATGCTTAAAGAAATACAGCCAAATCTCCAGTCCATATGAGAGGAATCTCAAACTGCAA 1249
DB 421 ATGCCCAGGAAATTTCTCCAACTTTGGTATCTTCTATAAAATCGGTGCACTGAGAATGCAG 362
QY 1250 TTCAGCAAAATGTGAACCGCTTAAAC 1274
DB 361 CTGGACCCCATGTGAAGATACCTTAAAC 337

RESULT 10
US-09-949-016-26415/c
; Sequence 26415, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
```

```

; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 26415
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-26415

Query Match
Best Local Similarity 61.2%; Score 99.8; DB 3; Length 601;
Matches 161; Conservative 0; Mismatches 102; Indels 0; Gaps 0;

QY 1012 AACTAAGGGGTGACTGCATACCGCTTGTCACAGGCTTGATGATGCAAGTGTGCGTGT 1071
DB 600 AAGTAGAGGAATCCATGCCAAGCGCTGTCCACAGGACTTTTCTGGTGAGCGCAGCTCT 541
QY 1072 GGAATCTGTGAAACGTGAGCTCTACGATTCGGCTTCTCTTAAGACTATCGAAGA 1131
DB 540 GGGTCTCTGTGAAGAGGTGGCCATCTATGGCTTCTGGCCCTTCTCTGTGAATATCATGA 481
QY 1132 CACCCCATCTAGTACCACTACTATGATAACATGTTTACCTAAGCATGGTTTCCACCAGAT 1191
DB 480 GCAGCCCATCAGCCACCACTACTATGACAAGCTCTTACCCCTTTCTGGCTTCCATGCCAT 421
QY 1192 GCCTAAAGAAATACAGCCAAATGCTCCAGCTCCATATGAGAGGAATCTCCAAACTGCAAT 1251
DB 420 GCCCGAGGAATTTCTCCAACTCTGGTATCTTTCATAAAATCGTGCACTGAGATGCAGCT 361
QY 1252 CAGCAAAATGTGAACGGCTTAACT 1274
DB 360 GGACCCATGTGAAGATACCTCAC 338

RESULT 11
US-09-949-016-72794/c
; Sequence 72794, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 72794
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-72794

Query Match
Best Local Similarity 61.2%; Score 99.8; DB 3; Length 601;
Matches 161; Conservative 0; Mismatches 102; Indels 0; Gaps 0;

QY 1012 AACTAAGGGGTGACTGCATACCGCTTGTCACAGGCTTGATGATGCAAGTGTGCGTGT 1071

```

```

DB 600 AAGTAGAGGAATCCATGCCAAGCGCTTGTCACAGGACTTTTCTGGTGAGCGCAGCTCT 541
QY 1072 GGAATCTGTGAAACGTGAGCTCTACGATTCGGCTTCTCTTAAGACTATCGAAGA 1131
DB 540 GGGTCTCTGTGAAGAGGTGGCCATCTATGGCTTCTGGCCCTTCTCTGTGAATATCATGA 481
QY 1132 CACCCCATCTAGTACCACTACTATGATAACATGTTTACCTAAGCATGGTTTCCACCAGAT 1191
DB 480 GCAGCCCATCAGCCACCACTACTATGACAAGCTCTTACCCCTTTCTGGCTTCCATGCCAT 421
QY 1192 GCCTAAAGAAATACAGCCAAATGCTCCAGCTCCATATGAGAGGAATCTCCAAACTGCAAT 1251
DB 420 GCCCGAGGAATTTCTCCAACTCTGGTATCTTTCATAAAATCGTGCACTGAGATGCAGCT 361
QY 1252 CAGCAAAATGTGAACGGCTTAACT 1274
DB 360 GGACCCATGTGAAGATACCTCAC 338

RESULT 12
US-08-626-994A-4
; Sequence 4, Application US/08626994A
; Patent No. 5798244
; GENERAL INFORMATION:
; APPLICANT: Shuichi TSUJI et al.
; TITLE OF INVENTION: Sialin, 2,3-Galactosyltransferase
; TITLE OF INVENTION: Sialin, 2,3-Galactosyltransferase
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Greenblum & Bernstein, P.L.C.
; STREET: 1941 Roland Clarke Place
; CITY: Reston
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 20191
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 1.44 diskette
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: Word Perfect 5.1+ (ASCII)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/626 994A
; FILING DATE: April 3, 1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 77469/1995
; FILING DATE: April 3, 1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Artold Turk
; REGISTRATION NUMBER: 33,094
; REFERENCE/DOCKET NUMBER: P145NS
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 716-1191
; TELEFAX: (703) 716-1180
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1048 bases
; TYPE: nucleic acid
; STRANDEDNESS: not relevant
; TOPOLOGY: linear
US-08-626-994A-4

Query Match
Best Local Similarity 50.2%; Score 79.6; DB 2; Length 1048;
Matches 283; Conservative 0; Mismatches 269; Indels 12; Gaps 3;

QY 559 CATCCCATTCGAGAGAACATTTCCACATGTTTCCAGTGTGGAGCTTTTGTGCGACTA 618
DB 301 CTTTCGATTAGCAATAACTTCGAGTCCCTGCTCCAGATGTGTGCCCATATGAATAA 360
QY 619 TCCCTATAACCAAGTGTGAGTGTGGTAAATGGGGGAATCTCAACAAGTCTCTCTGGGG 678
DB 361 CGTTTATAATGTTTGTGCTGTGGTGGAAACAGTGGATCTTGACAGGAGTCACTGTGG 420

```

QY 679 AGCAGAAATTGATAAATCTGACTCTCTCTCAGGTGTACCTCCCCCAATCACAGGGAG 738
DB 421 ACAAGAAATAGATAAATCAGATTTGTTCTCGATGCAATTTTGCCCC--GACAGAGGC 477
QY 739 CCGTAGTAAGATGTTGGAAGCAAAACAAATCTTGTGACTGTCAATCCAGCATTTATAAC 798
DB 478 TTTCCACAAAGATGTTGGAAGGAAACCAACCTCACAACTTCAATCCGAGCATCTTAGA 537
QY 799 CCGTAGTACCAAGATTT-----GAAGGAGAAAGAACAGATGTTTGGAGGACATCTC 852
DB 538 GAAATATTACAAATCTTTTAAACCATTCAGGACCGTAAACAATCTTCTCCTCAGTTTAAA 597
QY 853 CACCTATGAGATGATCTCTCTCTCTGAGCATTTTCTATCGGGCAACACAGGCAT 912
DB 598 AAGCTTATGAGGCGCATCTTTGGATCCCTGCAATTTTCTTCCACATCTTGCACATGT 657
QY 913 CTCTTTTAAAGTCTACCAACACTCAAGAGTCAAAAATGAGGCAAAAGGTTCTCTTCTT 972
DB 658 AACGAGAACGCTAGTGGATTTTGTGAGCACAGAGGTCAAGTTAAGGTCAGTTGGC 717
QY 973 CCATCCAGGTA---CCTGAGACCTCGCTCTTTCTGGAGAACTAAAGGGGTGACTGC 1029
DB 718 TTGGCTTGGAAATATCATGCAACATGTCAACAGGTACTTGGAAAAACAAACACCTGTACC 777
QY 1030 ATACCGCTTGTCCACAGCTTGTGATTCGAGTGTCTGGAAGTGTGTGAAAAAGT 1089
DB 778 CAACAGACTGACAGCAGGTATCTTAATGTATCTCTTGCATCTGCAATATGTGAAGAGAT 837
QY 1090 GAAGCTCTACGATTTCTGGCCCTT 1113
DB 838 CCACTTGTACGGTTCTGGCCCTT 861

RESULT 13

US-08-957-742-4
; Sequence 4, Application US/08957742
; Patent No. 6017743
; GENERAL INFORMATION:
; APPLICANT: Shuichi TSUJI et al.
; TITLE OF INVENTION: Sia' 2,3Gala 1,4G1cNAC ' 2,8-
; TITLE OF INVENTION: SIALYLTRANSFERASE
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Greenidium & Bernstein, P.L.C.
; STREET: 1941 Roland Clarke Place
; CITY: Reston
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 20191
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 1.44 diskette
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: Word Perfect 5.1+ (ASCII)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/957,742
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/626,994
; FILING DATE: April 3, 1996
; APPLICATION NUMBER: JP 77469/1995
; FILING DATE: April 3, 1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Arnold Turk
; REGISTRATION NUMBER: 33,094
; REFERENCE/DOCKET NUMBER: P4595
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 716-1191
; TELEFAX: (703) 716-1180
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:

; LENGTH: 1048 bases
; TYPE: nucleic acid
; STRANDEDNESS: not relevant
; TOPOLOGY: linear
US-08-957-742-4

Query Match 2.5%; Score 79.6; DB 3; Length 1048;
Best Local Similarity 50.2%; Prem. No. 5e-11; Mismatches 269; Indels 12; Gaps 3;

Matches 283; Conservative 0;
QY 559 CATCCCAATTCAGAGAAACATTTCCACATGTTTCCAGTGTCCAGCCCTTTTGTGACTA 618
DB 301 CTTCTCGATTAGCAATAACTTCCGTCCTGCTCCAGATGTGCGCCCATTTATGAATAA 360
QY 619 TCCCTATACCAAGTGTGAGTGTGTTAATGGGGAAATTTCAACAAGTCTCTCTGCGG 678
DB 361 CGCTTATAATGTTTGTGCTGTGTTGGAACACAGTGGAAATCTTGACAGGGAGTCAGTGTGG 420
QY 679 AGCAGAAATTCATAAATCTGACTCTCTCAGGTGTAAACCTCCCCCAATCACAGGGAG 738
DB 421 ACAAGAAATAGATAAATCAGATTTGTTCTCGATGCAATTTTGCCCC--GACAGAGGC 477
QY 739 CGCTAGTAAAGATGTTGGAAGCAAAACAAATCTTGTGACTGTCAATCCAGCATTTATAAC 798
DB 478 TTTCCACAAAGATGTTGGAAGCAAAACCAACCTCAACCTTCAATCCGAGCATCTTAGA 537
QY 799 CCGTAGTACCAAGATTT-----GAAGGAGAAAGAACAGATGTTTGGAGGACATCTC 852
DB 538 GAAATATTACAAATCTTTTAAACCATTCAGGACCGTAAACAATCTTCTCCTCAGTTTAAA 597
QY 853 CACCTATGAGATGATCTCTCTCTGAGCATTTTCTATCGGGCAACACAGGCAT 912
DB 598 AAGCTTATGAGGCGCATCTTTGGATCCCTGCAATTTTCTTCCACATCTTGCACATGT 657
QY 913 CTCTTTTAAAGTCTACCAACACTCAAGAGTCAAAAATGAGGCAAAAGGTTCTCTTCTT 972
DB 658 AACGAGAACGCTAGTGGATTTTGTGAGCACAGAGGTCAAGTTAAGGTCAGTTGGC 717
QY 973 CCATCCAGGTA---CCTGAGACCTCGCTCTTTCTGGAGAACTAAAGGGGTGACTGC 1029
DB 718 TTGGCTTGGAAATATCATGCAACATGTCAACAGGTACTTGGAAAAACAAACACCTGTACC 777
QY 1030 ATACCGCTTGTCCACAGCTTGTGATTCGAGTGTCTGGAAGTGTGTGAAAAAGT 1089
DB 778 CAACAGACTGACAGCAGGTATCTTAATGTATCTCTTGCATCTGCAATATGTGAAGAGAT 837
QY 1090 GAAGCTCTACGATTTCTGGCCCTT 1113
DB 838 CCACTTGTACGGTTCTGGCCCTT 861

RESULT 14

US-08-626-994A-2
; Sequence 2, Application US/08626994A
; Patent No. 5798244
; GENERAL INFORMATION:
; APPLICANT: Shuichi TSUJI et al.
; TITLE OF INVENTION: Sia' 2,3Gala 1,4G1cNAC ' 2,8-
; TITLE OF INVENTION: SIALYLTRANSFERASE
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Greenidium & Bernstein, P.L.C.
; STREET: 1941 Roland Clarke Place
; CITY: Reston
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 20191
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 1.44 diskette
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: Word Perfect 5.1+ (ASCII)
; CURRENT APPLICATION DATA:

Qy	1030	ATACCGCTTGTCCACAGGCTTGATGATTGCAAGTGTCCCTGTGGAACTGTGTGAAAACGT	1089
Db	962	CAAAACGACTGAGCACAGGTATCCTAATGTATATACTCTTGCACTCTGCAATATGTGAAGAT	1021
Qy	1090	GAAGCTCTACGGATTCTGGCCTTT	1113
Db	1022	CCACTTGTACGGTTTCTGGCCCTT	1045

Search completed: May 31, 2006, 14:25:31
 Job time : 572 secs